



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 966506

TO: Sheridan Swope
Location: cm1/12d12/10d01
Art Unit: 1652
Friday, June 20, 2003
Case Serial Number: 966880

From: Paul Schulwitz
Location: Biotech-Chem Library
CM1-6B06
Phone: 305-1954

paul.schulwitz@uspto.gov

Search Notes

Examiner Swope,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
STIC Biotech/Chem Library
(703)305-1954



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 11:07:40 ; Search time 6987 Seconds
(without alignments)
12781.153 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacatgctca.....tcaactctcagctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	272	4.9	293	17	AQ042682 CIT-HSP-2
2	87	1.6	693	12	BG757392 602711022
3	84	1.5	541	12	BF238155 601811880
4	84	1.5	743	12	BG686133 602638412
5	84	1.5	942	12	BF975166 602244657
6	71	1.3	953	14	BQ065440 AGENCOURT

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AQ042682	CIT-HSP-2326M11	293 bp DNA	AQ042682	1	GI:3310067	human.	Homo sapiens	1	Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M., and Venter, J.C.	Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)	Unpublished (1998)	Other GSSs: CIT-HSP-2326M11.TV
AQ042682	CIT-HSP-2326M11	293 bp DNA	AQ042682	1	GI:3310067	human.	Homo sapiens	1	Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M., and Venter, J.C.	Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)	Unpublished (1998)	Other GSSs: CIT-HSP-2326M11.TV

ALIGNMENTS

BQ055935 AGENCOURT
T06576 EST04465 Fe
AG056000 Pan. trogl
AM303243 Xr90C01.x
A1811647 tW4C03.x
BF086490 RC0-C1003
BE177917 RC3-HT060
AM562095 SMOYAFCAP
AA937809 nW8B02.s
BE169901 CM0-HT052
AA582746 nW39A11.s
AA569235 nm31a10.s
AA799144 RC0-UM005
AA224860 nC32G07.r
AM799150 RC0-UM005
AO559321 HS-5313_B
AO056100 CIT-HSP-2
AA993560 OT96G03.s
AO633793 RPT1-11-4
A1376179 tW59G08.x
AM799331 RC0-UM005
BE151195 RC4-HT027
AL601949 DKFP313D
BE151208 RC4-HT027
BF815902 MR2-C1012
AU147853 AU147853
AM026595 w45802.x
AV733437 AV733437
AO389816 RPT11-14
BG621190 602616889
AO379336 RPT11-16
BM993430 UI-H-DP0
AG156377 Pan. trogl
AO321493 RPT11-92
AG158392 Pan. trogl
AL704287 DKFP686P
AG178815 Pan. trogl
AG173021 Pan. trogl
AQ428052 CITBT-B1-

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends

FEATURES

source Location/Qualifiers
1..293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2326M11"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 69 a 74 c 91 g 59 t

Query Match 4.9%; Score 272 DB 17; Length 293;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3535 GACATTTAAAGTGAAGACAGACAGCCAGGTGTGGTGCACGCTGTAAATCCAGCAGCTC 3594
1 GACATTTAAAGTGAAGACAGACAGCCAGGTGTGGTGCACGCTGTAAATCCAGCAGCTC 60
QY 3595 TGGAGGCTGAGGTGGTGGATCGCTGAGCCCTGGAGTTCAAGACAGCCTTGAGACA 3654
61 TGGAGGCTGAGGTGGTGGATCGCTGAGCCCTGGAGTTCAAGACAGCCTTGAGACA 120
QY 3655 TGGCAAAACCTGTTTCTATACAAAATAGCCGGCATGTGGCATGTGCTGTGTC 3714
121 TGGCAAAACCTGTTTCTATACAAAATAGCCGGCATGTGGCATGTGCTGTGTC 180
QY 3715 CCAGCTACTAGGAGGCTGAGCAGGAAATCTTTGGAGCCAGAGCTCAAGCTGCAC 3774
181 CCAGCTACTAGGAGGCTGAGCAGGAAATCTTTGGAGCCAGAGCTCAAGCTGCAC 240
QY 3775 GAGCAGCTTGGCCGACCTGCACCTGCAGCCTG 3806
241 GAGCAGCTTGGCCGACCTGCACCTGCAGCCTG 272

RESULT 2 693 bp mRNA linear EST 15-MAY-2001
BG757392
LOCUS 602211022F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851580 5',
DEFINITION mRNA sequence.
ACCESSION BG757392
VERSION BG757392.1 GI:14068045
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov

FEATURES

source Location/Qualifiers
1..693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4851580"

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1694 row: k column: 05
High quality sequence stop: 693.

/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOT87; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(6). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: This is a NIH_MGC Library."

BASE COUNT 157 a 188 c 178 g 170 t

Query Match 1.6%; Score 87; DB 12; Length 693;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAGAACATCATTAATGAGATGATTTTCTGCTGAGACTTGCAGGAGGAGCAAGA 1091
1 AGAGAACATCATTAATGAGATGATTTTCTGCTGAGACTTGCAGGAGGAGCAAGA 61
QY 1092 AGACACTGTGACACCACTATGACAG 1118
62 AGACACTGTGACACCACTATGACAG 88
Db

RESULT 3 541 bp mRNA linear EST 14-NOV-2000
BF238155
LOCUS 601811880F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054515 5',
DEFINITION mRNA sequence.
ACCESSION BF238155
VERSION BF238155.1 GI:11152074
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM895 row: p column: 20
High quality sequence stop: 541.

FEATURES

source Location/Qualifiers
1..541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4054515"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOT87; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(6). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: This is a NIH_MGC Library."

BASE COUNT 120 a 150 c 142 g 129 t

```

ORIGIN
Query Match      1.5%; Score 84; DB 12; Length 541;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1035 GAACCATCATTAATTAAGTGAAGATTTTCTGGCGCTGAGACTTCGAGGAGGCAAGAAGA 1094
         |||||||
Db      2 GAACCATCATTAATTAAGTGAAGATTTTCTGGCGCTGAGACTTCGAGGAGGCAAGAAGA 61
         |||||||

Oy      1095 CACTCTGGACACCACTATGACAG 1118
         |||||||
Db      62 CACTCTGGACACCACTATGACAG 85
         |||||||

RESULT 4
BG686133      743 bp mRNA linear EST 01-MAY-2001
LOCUS      602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5',
DEFINITION mRNA sequence.
ACCESSION BG686133
VERSION BG686133.1 GI:13917530
KEYWORDS EST.
SOURCE EST.
ORGANISM human.
          Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: LNCM1626 row: 9 column: 03
          High quality sequence stop: 740.
          Location/Qualifiers
            1..743
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4766234"
              /clone_lib="NIH_MGC_48"
              /tissue_type="primary B-cells from tonsils (cell line)"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
              Site_2: EcoRI; cDNA made by oligo-dT priming.
              Directionally cloned into EcoRI/XhoI sites using the
              following 5' adaptor: GGCACGAG(G). Size-selected >500bp
              for average insert size 1.8kb. Library constructed by Ling
              Hong in the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies).
              Note: this is a NIH_MGC Library."
BASE COUNT      176 a      197 c      188 g      182 t
ORIGIN
Query Match      1.5%; Score 84; DB 12; Length 743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1035 GAACCATCATTAATTAAGTGAAGATTTTCTGGCGCTGAGACTTCGAGGAGGCAAGAAGA 1094
         |||||||
Db      2 GAACCATCATTAATTAAGTGAAGATTTTCTGGCGCTGAGACTTCGAGGAGGCAAGAAGA 61
         |||||||

Oy      1095 CACTCTGGACACCACTATGACAG 1118
         |||||||
Db      62 CACTCTGGACACCACTATGACAG 85
         |||||||

```

```

RESULT 5
BF975166      942 bp mRNA linear EST 22-JAN-2001
LOCUS      602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4335639 5',
DEFINITION mRNA sequence.
ACCESSION BF975166
VERSION BF975166.1 GI:12342381
KEYWORDS EST.
SOURCE EST.
ORGANISM human.
          Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: LNCM1207 row: a column: 16
          High quality sequence stop: 707.
          Location/Qualifiers
            1..942
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4335639"
              /clone_lib="NIH_MGC_48"
              /tissue_type="primary B-cells from tonsils (cell line)"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
              Site_2: EcoRI; cDNA made by oligo-dT priming.
              Directionally cloned into EcoRI/XhoI sites using the
              following 5' adaptor: GGCACGAG(G). Size-selected >500bp
              for average insert size 1.8kb. Library constructed by Ling
              Hong in the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies).
              Note: this is a NIH_MGC Library."
BASE COUNT      249 a      209 c      251 g      233 t
ORIGIN
Query Match      1.5%; Score 84; DB 12; Length 942;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1035 GAACCATCATTAATTAAGTGAAGATTTTCTGGCGCTGAGACTTCGAGGAGGCAAGAAGA 1094
         |||||||
Db      2 GAACCATCATTAATTAAGTGAAGATTTTCTGGCGCTGAGACTTCGAGGAGGCAAGAAGA 61
         |||||||

Oy      1095 CACTCTGGACACCACTATGACAG 1118
         |||||||
Db      62 CACTCTGGACACCACTATGACAG 85
         |||||||

RESULT 6
BQ065440      953 bp mRNA linear EST 02-APR-2002
LOCUS      AGENCOURT 6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
DEFINITION 5', mRNA sequence.
ACCESSION BQ065440
VERSION BQ065440.1 GI:19894486
KEYWORDS EST.
SOURCE EST.
ORGANISM human.
          Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(Pages 1 to 933)	Mammalia: Eutheria: Primates; Catarrhini; Hominoidea: Homo.		
	NIH-MGC	http://mgc.ncl.nih.gov/ .		
	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)		
	Contact: Robert Strusberg, Ph.D.			
	Email: cgabbs-r@mail.nih.gov			
	Tissue Procurement: Lou Staudt			
	CDNA Library Preparation: Rubin Laboratory			
	DNA Sequencing by: The I.M.A.G.E. Consortium (LINL)			
	Clone distribution by: Agencourt Bioscience Corporation			
	found through the I.M.A.G.E. Consortium/LINL at:			
	http://image.lnl.gov			
	Place: LRCM2108	row: p	column: 10	
	High quality sequence stop: 634.			
	Location/Qualifiers			
	1..933			

```

FEATURES
Source
1. .933
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5929377"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCR7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." Note: this is a NIH_MGC library."
238 a 236 c 233 g 246 t
BASE COUNT
ORIGIN

```

Query Match	1.38;	Score 71;	DB 14;	Length 953,
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	71;			

	TTGAGTGA	TATTTCTCCCGC	TAATGAA	0;	Gaps	0;
QY	1048					

DB 1 TTGAAGTGAAGATTTTCTCGGCGCTGAGACTTGCAGGGGAGGCGAAGAAGACACTCTGGACACC 1107
 QY 1108 ACTATGAGACAG 1118
 |||||

61 ACTATGCACAG 71

RESULT		7
B005935		
LOCUS		
DEFINITION	B0055935	1052 bp mRNA linear EST 29-MAR-2002
ACCESSION	AB06NCOURT_6796291	NH_MGC_99 Homo sapiens cDNA clone IMAGE:5808181
VERSION	B0055935	
KEYWORDS	B0055935.1	GI:19815262
SOURCE	EST.	
ORGANISM	human.	

CONVULSISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS
TITLE
NIM-MGC <http://mgc.ncl.nih.gov/>.
JOURNAL
National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT
Unpublished (1999)
Contact: Robert S. Stein

Small: cgsphs@small.nth.se, Ph. D.
Tissue procurement: all.nth.gov
CDNA Library Preparation: Staudt
CDNA Library Arrived by: I.M.A.G.E. Consortium
DNA Sequencing by: Agilent BioScience Corporation
Clone distribution: MCC clone distribution
found through the I.M.A.G.E. Consortium/LLNL at:

```

http://image.llnl.gov
plate: L1CM2051 row: m column: 14
High quality sequence stop: 665.
location/Qualifiers
1..1052
FEATURES
source
```

```

BASE COUNT
ORIGIN
276 a 250 c 258 g 266 t 2 others
//samsum="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5808181"
/clone_1ib="NIH_MGC_99"
/tissue_type="Lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOT7; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
1.8kb. library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
11 Rn (Life Technologies). Note: this is a NIH_MGC
library."

```

Query Match	1.38;	Score 71;	DB 14;	Length 1052;
Best Local Similarity	100.08;	Pred. No. 0;		
Matches 71;	Conservative	0.00		

		ends	0;	Gaps	0;
QY	1048	TTGAAGTGAGATTTTCTGCGCCTGAGACTTCACCCACCAATT			

Db 1 TTGAAGTAGAATTTTCTTGCGCTTGAGACTTGACGGGAGCGAAGAGACACTTGGACACC 60
QY 1108 ACTATGACAG 1118

61 ACTATGGACAG 71

RESULT 8	LOCUS	DEFINITION	Accession	Version	Keywords	Source	Organism
T06576	T06576	EST004465 Fetal brain. Stragagene (cat#36206) Homo sapiens CDNA clone HEBDV37 similar to EST containing Alu repeat.	423 bp	linear	EST 30-TUN-1993	human.	Homo sapiens

REFERENCE
1 (bases 1 to 423)
Adams, M. D. *Journal of Molecular Evolution* 31: 1-10 (1990)

TITLE 3,400 expressed sequence tags identify diversity of transcripts from human brain
JOURNAL Nat. Genet. 4, 256-267 (1993)
MEDLINE 93364420
COMMENT Contact: Adams, MD

FEATURES
 source 1. 423
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878
 Tel: 3018609080
 Fax: 3018699473
 Email: mdadams@tigr.org
 Seq primer: M13-71
 Location/Qualifiers
 1. 423

```

/organism="Homo sapiens"
/db_xref="ATCC (InDSTL):83234"
/db_xref="taxon:9606"
/clone="HERDV37"
/clone_lip="Fetal brain, Striatogene (cat#936206)"
/note="Vector: lambdaZAP-IT: 17-18 wk gestation, fema
vector, 1.0kb average cDNA syntheses; lambdaZAP-IT
BASE COUNT          95 a      109 c      96 g      117 t      6 others
ORIGIN

```

Fri Jun 20 09:21:57 2003

us-09-966-880a-9.oli.rst

```

RESULT 9
AG056000/c
LOCUS      AG056000
DEFINITION Pan troglodytes DNA, clone: PTB-042C11.R, genomic survey sequence.
ACCESSION  AG056000
VERSION     AG056000.1 GI:16593459
KEYWORDS
SOURCE
ORGANISM   Pan troglodytes male lymphoblast DNA, clone: PTB Chimpanzee Male
           BAC library clone: PTB-042C11.R.
           BAC library clone: PTB-042C11.R.
           Pan troglodytes male lymphoblast DNA, clone: PTB Chimpanzee Male
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.

REFERENCE  1
AUTHORS   Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
           Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE     Unpublished
JOURNAL   2 (bases 1 to 648)
AUTHORS   Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
           Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE     Direct Submission
JOURNAL   Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical
           and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
           Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB this BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.
           PRIMERS
           Sequencing: M13Rev
           LIBRARY
           Vector : pKS145
           R.site 1 : SacI
           R.site 2 : SacI.
           Location/Qualifiers
           1. 648
           /organism="Pan troglodytes"
           /db_xref="taxon:9598"
           /clone="PTB-042C11.R"
           /sex="male"
           /cell_type="lymphoblast"
           /clone_id="PTB Chimpanzee Male BAC library"
           174 a 144 c 177 g 153 t
BASE COUNT 174 a 144 c 177 g 153 t
ORIGIN
Query Match 1.3%; Score 69; DB 17; Length 648;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5439 GCTAATGTTTGTATTTAGTAGAGATGGGCTTCCACCATGTTGGCCAGGCTGCTCA 5498
DB 558 GCTAATGTTTGTATTTAGTAGAGATGGGCTTCCACCATGTTGGCCAGGCTGCTCA 499
OY 5499 ACTGCTGAC 5507
DB 498 ACTGCTGAC 490

RESULT 10
AM303243/c
LOCUS      AM303243
DEFINITION x190c01.x1 NCI CGAP Lu26 Homo sapiens cDNA clone IMAGE:2767392.3'
           similar to contains Alu repetitive element,, mRNA sequence.
ACCESSION  AM303243
VERSION     AM303243.1 GI:6712923
KEYWORDS
SOURCE
ORGANISM   human.
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  1 (bases 1 to 429)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
           Unpublished (1997)
           Contact: Robert Strausberg, Ph.D.
           Email: cgaps-r@mail.nih.gov
           cDNA Library Preparation: David B. Krizman, Ph.D.
           cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/biopr/image/image.html
           Seq primer: -400p from Gibco
           High quality sequence stop: 424.
           Location/Qualifiers
           1. 429
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="IMAGE:2767392"
           /clone_id="NCI CGAP_Lu26"
           /tissue_type="invasive adenocarcinoma"
           /dev_stage="adult"
           /lab_host="DH10B"
           /note="Organ: Lung; Vector: PAMPI; mRNA made from lung
           adenocarcinoma tissue. cDNA made by oligo-dT priming,
           directionally cloned. Size-selected on agarose gel,
           average insert size 500 bp. Primary library,
           non-amplified."
           non-amplified.
           138 a 80 c 110 g 101 t
BASE COUNT 138 a 80 c 110 g 101 t
ORIGIN
Query Match 1.2%; Score 68; DB 10; Length 429;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5446 TTGTGATTTTGTAGAGATGGGCTTCCACCATGTTGGCCAGGCTGCTCAACTCTG 5505
DB 349 TTGTGATTTTGTAGAGATGGGCTTCCACCATGTTGGCCAGGCTGCTCAACTCTG 290
OY 5506 ACTCTGAC 5513
DB 289 ACTCTGAC 282

RESULT 11
A1811647
LOCUS      A1811647
DEFINITION t44c03.x1 NCI CGAP Utl Homo sapiens cDNA clone IMAGE:2262532.3'
           similar to contains Alu repetitive element,, mRNA sequence.
ACCESSION  A1811647
VERSION     A1811647.1 GI:5398213
KEYWORDS
SOURCE
ORGANISM   human.
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  1 (bases 1 to 578)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index

```

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-r@emall.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/RLNL at:
www.dio.llnl.gov/dbcrp/image/image.html
Insert length: 1045 Std Error: 0.00
Seq primer: -400P from Glibco
High quality sequence stop: 406.
Location/Qualifiers

FEATURES
source

```
1. 578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Ut1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 Kb. Life Technologies catalog #:
11538-014"
```

BASE COUNT 147 a 139 c 114 g 177 t 1 others

ORIGIN

Query Match 1.28; Score 68; DB 9; Length 578;
Best Local Similarly 100.0%; Pred. No. 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5446 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGTCACAACTCTG 5505
|||||
Db 130 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGTCACAACTCTG 189
OY 5506 ACCTCAGA 5513
|||||
Db 190 ACCTCAGA 197

RESULT 12
LOCUS

BF806490 164 bp mRNA linear EST 12-JAN-2001
R00-CI0037-081100-032-d01_1 CI0037 Homo sapiens cDNA, mRNA

ACCESSION
VERSION

BF806490
EST.
BF806490.1 GI:12135479

KEYWORDS
SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 164)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL
MEDLINE
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC06t2=RC0-CI0037-
081100-032-d01_1&t3=2000-11-08&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 164.
Location/Qualifiers

FEATURES
source

```
1. 164
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0037"
/dev_stage="Adult"
/Note="Organ: colon; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORBESTES PCR (U.S. Letters Patent application
No. 196/716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

BASE COUNT 41 a 36 c 52 g 35 t

ORIGIN

Query Match 1.28; Score 67; DB 12; Length 164;
Best Local Similarly 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5446 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGTCACAACTCTG 5505
|||||
Db 79 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGTCACAACTCTG 20
OY 5506 ACCTCAG 5512
|||||
Db 19 ACCTCAG 13

RESULT 13

BE177917 198 bp mRNA linear EST 22-JUN-2000
RC3-HT0600-170300-011-g01_1 HT0600 Homo sapiens cDNA, mRNA

LOCUS

BE177917
sequence.
BE177917
KEYWORDS
EST.
BE177917.1 GI:8657069

VERSION
SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 198)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL
MEDLINE
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3-HT0600-170
300-011-g01_1&t3=2000-03-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 198.

FEATURES
source
Location/Qualifiers
1. 198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0600"
/dev_stage="Adult"
/note="Organ: head, neck; Vector: puc18; site_1: Sma1; site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT
55 a 43 c 59 g 41 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1.2%; Score 67; DB 10; Length 198;
OY 5446 TTGTGATTTTGTAGATGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 5505
|||||
Db 79 TTGTGATTTTGTAGATGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 20
|||||
OY 5506 ACCTCAG 5512
|||||
Db 19 ACCTCAG 13
|||||
RESULT 14
AM562095 256 bp mRNA linear EST 09-MAR-2000
LOCUS AM562095.1
DEFINITION SMOVAFCAP34C01SK Onchocerca volvulus adult female cDNA (SAM98MLM-OVAF) Onchocerca volvulus cDNA clone SMOVAFCAP34C01 5', mRNA sequence.
ACCESSION AM562095
VERSION AM562095.1 GI:7212211
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 256)
AUTHORS Lizotte-Maniowski, M. and Williams, S.A.
TITLE Genes expressed in adult female stage of Onchocerca volvulus
JOURNAL Unpublished (1998)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pbluescript SK.
FEATURES
source
Location/Qualifiers
1. 256
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone_lib="SMOVAFCAP34C01"
/clone_lib="Onchocerca volvulus adult female cDNA (SAM98MLM-OVAF)"
/sex="female"
/dev_stage="adult"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda uni-zap XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Two adult female worms of Onchocerca volvulus were isolated from consenting patients and quick frozen. Adult female mRNA was converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7 x 10E5 independent recombinants and the average insert size is ~1100bp. The library was

constructed by Michelle Lizotte-Maniowski with worms provided by Dr. Sara Lustigman. The library is available from Dr. Steven A. Williams, email: genomesmith.edu."
BASE COUNT
68 a 60 c 80 g 48 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1.2%; Score 67; DB 10; Length 256;
OY 5446 TTGTGATTTTGTAGATGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 5505
|||||
Db 98 TTGTGATTTTGTAGATGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 39
|||||
OY 5506 ACCTCAG 5512
|||||
Db 38 ACCTCAG 32
|||||
RESULT 15
AA937809 300 bp mRNA linear EST 30-APR-1998
LOCUS AA937809
DEFINITION nw89b02.s1 NCI-CGAP_P12 Homo sapiens cDNA clone IMAGE:125373 similar to contains Alu repetitive element; contains element L1 repetitive element; mRNA sequence.
ACCESSION AA937809
VERSION AA937809.1 GI:3095920
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 300)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbcr@mail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/HLN at: www.bio.lnhi.gov/birp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES
source
Location/Qualifiers
1. 300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1253739"
/clone_lib="NCI-CGAP_P12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
BASE COUNT
55 a 84 c 67 g 94 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1.2%; Score 67; DB 9; Length 300;
OY 5446 TTGTGATTTTGTAGATGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 5505
|||||
Db 168 TTGTGATTTTGTAGATGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 227
|||||

Fri Jun 20 09:21:57 2003

us-09-966-880a-9.oli.rst

Oy	5506	ACCTCAG	5512
Db	228	ACCTCAG	234

Search completed: June 19, 2003, 13:52:28
Job time : 6992 secs

Alignment Scores:

Pred. No.: 0.005 Length: 169
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.69% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x S57138 (1-169)

QY 2840 ACACACACACACACACACACACACACACACAC 2875

Db 18 ThriHsthrHsthrHsthrHsthrHsthrHsthrHst 29

RESULT 7

artifact-warning sequence (translated ALU class B) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: B40201

R:Claverie, J.M.

Personal communication, 1992

A:Reference number: A40201

A:Accession: B40201

A:Molecule type: DNA

A:Residues: 1-301 <CLAA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'x'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.: 0.00486 Length: 301
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.69% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x B40201 (1-301)

QY 3614 CCACCCACCTCAGCCTCCAGAGTGTGGATTACA 3579

Db 283 ProProthrsrAlasrGlnserAlaglylththr 294

RESULT 8

artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

Personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 <CLAA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'x'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.: 0.0047 Length: 613
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.69% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x C40201 (1-613)

QY 3568 TGCTCAGCCTGTAATCCAGACACTGTGGAGGCT 3603

Db 5 TrrpleuthrProValleProAlaleutrrpAla 16

RESULT 9

artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

Personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 <CLAA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o

in-frame stop codons are shown as 'x'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:

Pred. No.: 0.0047 Length: 613
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.69% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x C40201 (1-613)

QY 3602 GCCTCCAGAGTGTGGATTACAGCGCTGAGCCAC 3567

Db 599 AlasrGlnserAlaglylththrGlyValserHs 610

RESULT 10

hypothetical protein DKFZp434G1035.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000

C:Accession: T42696

R:Postula, A.; Klein, M.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, November 1999

A:Reference number: 222232

A:Accession: T42696

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-163 <AAA>

A:Cross-references: EMBL:AL133039

A:Experimental source: adult testis; clone DKFZp434G1035

C:Genetics:

A:Note: DKFZp434G1035.1

Alignment Scores:

Pred. No.: 0.0573 Length: 163
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.63% Indels: 0

Db 442 LeuLysLysIleArgArgLysIleLysAsnLys 452

RESULT 15

E40201

artifact-warning sequence (translated Alu class E) - human

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: E40201

R:Clavertie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: E40201

A:Molecule type: DNA

A:Residues: 1-597 <CLA>

R:Clavertie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.:	0.0537	Length:	597
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.63%	Indels:	0
DB:	4	Gaps:	0

US-09-966-880A-9 (1-5514) x E40201 (1-597)

QY 3558 CCAGGTGTGTGGCTCAGCCTGTATCCAGC 3590
 |||||
 Db 201 ProGlyValValAlaHisAlaCysAsnProSer 211

Search completed: June 14, 2003, 19:35:41
 Job time : 181.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 07:46:55 ; Search time 216 Seconds
(without alignments)
7828.776 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514

Sequence: 1 acagacgaatacatagtcca.....tcaaacctctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	1.2	20303	1 US-08-370-975B-6	Sequence 6, Appli
2	67	1.2	26764	1 US-08-370-975B-1	Sequence 1, Appli
3	66	1.2	8758	4 US-09-799-345-3	Sequence 3, Appli
4	63	1.1	99500	4 US-09-798-096-10	Sequence 10, Appl
5	61	1.1	2559	2 US-08-724-774B-3	Sequence 3, Appli
6	61	1.1	2559	3 US-09-089-595-3	Sequence 3, Appli
7	61	1.1	2559	4 US-09-382-855-3	Sequence 3, Appli
8	61	1.1	2559	4 US-09-183-714B-3	Sequence 3, Appli
9	61	1.1	2559	4 US-09-642-281-3	Sequence 3, Appli
10	61	1.1	43950	4 US-09-735-934A-3	Sequence 3, Appli
11	60	1.1	1541	6 5183884-1	Patent No. 5183884
12	60	1.1	1542	1 US-07-978-895-1	Sequence 1, Appli
13	60	1.1	1542	1 US-08-473-119-1	Sequence 1, Appli
14	60	1.1	1542	2 US-08-475-352-1	Sequence 3, Appli
15	58	1.1	36159	4 US-09-749-588-3	Sequence 382, App
16	56	1.0	57	1 US-08-222-177A-382	Sequence 49, Appl
17	56	1.0	4576	2 US-08-832-877-49	Sequence 3, Appli
18	55	1.0	4576	2 US-08-301-665-3	Sequence 3, Appli
19	55	1.0	36741	4 US-09-738-894A-3	Sequence 3, Appli
20	54	1.0	36651	4 US-09-738-894A-3	Sequence 3, Appli
21	54	1.0	48974	4 US-08-920-422-17	Sequence 17, Appl
22	54	1.0	112132	4 US-09-741-150-3	Sequence 3, Appli
23	52	0.9	3474	4 US-09-527-236A-1	Sequence 1, Appli
24	52	0.9	6140	4 US-09-605-785-536	Sequence 536, App
25	52	0.9	6140	4 US-09-439-313-536	Sequence 536, App
26	51	0.9	29629	4 US-09-729-995-3	Sequence 3, Appli
27	51	0.9	59065	4 US-09-813-817-3	Sequence 3, Appli

c 28 51 0.9 59065 4 US-09-978-197-3 Sequence 3, Appli
c 29 51 0.9 81001 4 US-09-750-580-1 Sequence 1, Appli
c 30 51 0.9 98844 4 US-09-791-211-10 Sequence 10, Appl
c 31 51 0.9 152331 3 US-09-128-155-16 Sequence 16, Appl
c 32 51 0.9 162450 4 US-09-345-882-1 Sequence 1, Appli
c 33 51 0.9 162450 4 US-09-128-155-17 Sequence 17, Appli
c 34 51 0.9 176373 3 US-09-128-155-17 Sequence 17, Appli
c 35 50 0.9 807 2 US-08-531-927B-9 Sequence 9, Appli
c 36 50 0.9 1776 2 US-08-531-927B-1 Sequence 12, Appli
c 37 50 0.9 1776 4 US-09-041-886-12 Sequence 20, Appli
c 38 50 0.9 6769 1 US-08-480-784-20 Sequence 20, Appli
c 39 50 0.9 6769 1 US-08-483-553-20 Sequence 20, Appli
c 40 50 0.9 6769 1 US-08-487-002-20 Sequence 20, Appli
c 41 50 0.9 6769 1 US-08-483-554B-20 Sequence 20, Appli
c 42 50 0.9 6769 1 US-08-488-011B-20 Sequence 20, Appli
c 43 50 0.9 6769 4 US-08-850-727-20 Sequence 20, Appli
c 44 50 0.9 6769 5 PCT-US95-10202-20 Sequence 20, Appli
c 45 50 0.9 6769 5 PCT-US95-10203-20 Sequence 20, Appli

ALIGNMENTS

RESULT 1
US-08-370-975B-6
; Sequence 6, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timlan, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-6

Query Match 1.2%; Score 67; DB 1; Length 20303;
Best Local Similarity 100.0%; Pred. No. 2.6e-16;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGTCTCAAACTCTG 5505
|||||

Db 6392 TTGTATTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 6451
 QY 5506 ACCTCAG 5512
 Db 6452 ACCTCAG 6458

RESULT 2

US-08-370-975B-1
 ; Sequence 1, Application US/08370975B
 ; Patent No. 5622851
 ; GENERAL INFORMATION:
 ; APPLICANT: Maley, Frank
 ; APPLICANT: Maley, Gladys F.
 ; APPLICANT: Weiner, Karen X.B.
 ; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Nixon, Hargrave, Devans & Doyle
 ; STREET: Clinton Square, P.O. Box 1051
 ; CITY: Rochester
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 14603

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,975B
 ; FILING DATE: 10-JAN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Timian, Susan J.
 ; REGISTRATION NUMBER: 34,103
 ; REFERENCE/DOCKET NUMBER: 20894/80
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (716)263-1636
 ; TELEFAX: (716)263-1600
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 26764 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; POSITION IN GENOME:
 ; CHROMOSOME/SEGMENT: 4q35
 ; US-08-370-975B-1

Query Match 1.2%; Score 67; DB 1; Length 26764;
 Best Local Similarity 100.0%; Pred. No. 2.5e-16;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTATTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 Db 8355 TTGTATTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 8414
 QY 5506 ACCTCAG 5512
 Db 8415 ACCTCAG 8421

RESULT 3

US-09-799-345-3
 ; Sequence 3, Application US/09799345
 ; Patent No. 6323016
 ; GENERAL INFORMATION:
 ; APPLICANT: YE, Jane et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL001156
 ; CURRENT APPLICATION NUMBER: US/09/799,345
 ; CURRENT FILING DATE: 2001-03-06
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 8758
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-799-345-3

Query Match 1.2%; Score 66; DB 4; Length 8758;
 Best Local Similarity 100.0%; Pred. No. 6.9e-16;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTATTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 Db 6196 TTGTATTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 6255
 QY 5506 ACCTCA 5511
 Db 6256 ACCTCA 6261

RESULT 4

US-09-798-096-10/c
 ; Sequence 10, Application US/09798096
 ; Patent No. 6399378
 ; GENERAL INFORMATION:
 ; APPLICANT: Donna T. Ward
 ; APPLICANT: Andrew T. Watt
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
 ; FILE REFERENCE: RTS-0207
 ; CURRENT APPLICATION NUMBER: US/09/798,096
 ; CURRENT FILING DATE: 2001-03-01
 ; NUMBER OF SEQ ID NOS: 89
 ; SEQ ID NO 10
 ; LENGTH: 99500
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; US-09-798-096-10

Query Match 1.1%; Score 63; DB 4; Length 99500;
 Best Local Similarity 100.0%; Pred. No. 5.9e-15;
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5450 TATTTTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5509
 Db 65234 TATTTTATTTAGATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 65175
 QY 5510 CAG 5512
 Db 65174 CAG 65172

RESULT 5

US-08-724-774B-3
 ; Sequence 3, Application US/08724774B
 ; Patent No. 5908778
 ; GENERAL INFORMATION:
 ; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coullie,
 ; APPLICANT: Pierre; Cerrozzini, Jean-Charles; Carrel,
 ; APPLICANT: Stefan; Reed, Daryl
 ; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor
 ; TITLE OF INVENTION: Rejection Antigen Precursors Mage-10,
 ; TITLE OF INVENTION: Antibodies Specific To The Molecule, and
 ; TITLE OF INVENTION: Uses Thereof
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felfe & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City

STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,774B
FILING DATE: 03-October-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5908778man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2559 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-774B-3

Query Match 1.1% Score 61; DB 2: Length 2559;
Best Local Similarity 100.0%; Pred. No. 6.1e-14;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5452 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
|||||
Db 2202 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261
Qy 5512 G 5512
|
Db 2262 G 2262

RESULT 6
US-09-089-595-3
Sequence 3, Application US/09089595
Patent No. 6153728
GENERAL INFORMATION:
APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coullie,
APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,
APPLICANT: Stefan; Reed, Daryl
TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor
TITLE OF INVENTION: Rejection Antigen Precursors Mage-10,
TITLE OF INVENTION: Antibodies Specific To The Molecule, and
TITLE OF INVENTION: Uses Thereof
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/089,595
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,774
FILING DATE: 03-October-1996
ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 6153728man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2559 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-089-595-3

Query Match 1.1% Score 61; DB 3: Length 2559;
Best Local Similarity 100.0%; Pred. No. 6.1e-14;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5452 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
|||||
Db 2202 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261
Qy 5512 G 5512
|
Db 2262 G 2262

RESULT 7

US-09-382-855-3
Sequence 3, Application US/09382855
Patent No. 6174692
GENERAL INFORMATION:

APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coullie, Pierre;
APPLICANT: Cerrottini, Jean-Charles; Carrel, Stefan; Reed, Daryl
TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor Rejection
TITLE OF INVENTION: Antigen Precursors Mage-10, Antibodies Specific To The Molecu
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski LLP
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/382,855
FILING DATE: 25-August-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/089,595
FILING DATE: 02-June-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,774
FILING DATE: 03-October-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Schofield, Mary Anne
REGISTRATION NUMBER: 36,669
REFERENCE/DOCKET NUMBER: LUD 5457.2 DIV - JEL/MAS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2559 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-382-855-3

```
Query Match
Best Local Similarity 1.1%; Score 61; DB 4; Length 2559;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5452 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
DB 2202 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261

QY 5512 G 5512
DB 2262 G 2262

RESULT 8
US-09-183-714B-3
; Sequence 3, Application US/09183714B
; Patent No. 6221593
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry
; APPLICANT: Brasseur, Francis
; APPLICANT: Rimoldi, Donata
; APPLICANT: De Plaen, Etienne
; TITLE OF INVENTION: Method for Determining Cancer by Determining Expression
; TITLE OF INVENTION: Of MAGE-10
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/183,714B
; CURRENT FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 08/724,774
; PRIOR FILING DATE: 1996-10-03
; NUMBER OF SEQ ID NOS: 7
; SEQ ID NO 3
; LENGTH: 2559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-183-714B-3
```

```
Query Match
Best Local Similarity 1.1%; Score 61; DB 4; Length 2559;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5452 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
DB 2202 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261

QY 5512 G 5512
DB 2262 G 2262
```

```
RESULT 9
US-09-642-281-3
; Sequence 3, Application US/09642281
; Patent No. 6387698
; GENERAL INFORMATION:
; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie, Pierre;
; Cerrotti, Jean-Charles; Carrel, Stefan; Reed, Daryl
; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor Rejection
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/642,281
; FILING DATE: 18-Aug-2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/089,595
; FILING DATE: 02-June-1998
; APPLICATION NUMBER: US 08/724,774
; FILING DATE: 03-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Schofield, Mary Anne
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5457.2 DIV - JEL/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2559 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-642-281-3
```

```
Query Match
Best Local Similarity 1.1%; Score 61; DB 4; Length 2559;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5452 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
DB 2202 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261

QY 5512 G 5512
DB 2262 G 2262
```

```
RESULT 10
US-09-735-934A-3/C
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: Li, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3
```

```
Query Match
Best Local Similarity 1.1%; Score 61; DB 4; Length 43950;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5450 TATTTTGTAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCT 5509
DB 35406 TATTTTGTAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCT 35347

QY 5510 C 5510
DB 35346 C 35346
```

```
RESULT 11
5183884-1
```

Patent No. 5183884
; APPLICANT: KRAUS, MATTHIAS H.; AARONSON, STUART A.
; TITLE OF INVENTION: DNA SEGMENT ENCODING A GENE FOR A
; RECEPTOR RELATED TO THE EPIDERMAL GROWTH FACTOR RECEPTOR
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/444,406
; FILING DATE: 01-DEC-1989
; SEQ ID NO:1:
; LENGTH: 1541
5183884-1

Query Match 1.1%; Score 60; DB 6; Length 1541;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
|||||
Db 1410 TTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1469
|||||

RESULT 12
US-07-978-895-1
; Sequence 1, Application US/07978895
; Patent No. 5480968
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREO, AND
; TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/978,895
; FILING DATE: 19921110
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 66..221
; FEATURE:
; NAME/KEY: exon
; LOCATION: 780..855
; FEATURE:
; NAME/KEY: exon

; LOCATION: 1040..1185
; FEATURE:
; NAME/KEY: intion
; LOCATION: 222..779
; FEATURE:
; NAME/KEY: intron
; LOCATION: 856..1039
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(66..221, 780..855, 1040..1185)
US-07-978-895-1

Query Match 1.1%; Score 60; DB 1; Length 1542;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
|||||
Db 1411 TTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1470
|||||

RESULT 13
US-08-473-119-1
; Sequence 1, Application US/08473119
; Patent No. 5820859
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREO, AND
; TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,119
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/978,895
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 66..221
; FEATURE:
; NAME/KEY: exon
; LOCATION: 780..855

```

; FEATURE:
; NAME/KEY: exon
; LOCATION: 1040..1185
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 222..779
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 856..1039
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(66..221, 780..855, 1040..1185)
US-08-473-119-1

Query Match
Best Local Similarity 1.1%; Score 60; DB 1; Length 1542;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
Db 1411 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1470

RESULT 14
US-08-475-352-1
; Sequence 1, Application US/08475352
; Patent No. 5916755
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREO, AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.352
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/978,895
; FILING DATE:
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 66..221
; FEATURE:

```

```

; NAME/KEY: exon
; LOCATION: 780..855
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1040..1185
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 222..779
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 856..1039
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(66..221, 780..855, 1040..1185)
US-08-475-352-1

Query Match
Best Local Similarity 1.1%; Score 60; DB 2; Length 1542;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
Db 1411 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1470

RESULT 15
US-09-749-588-3
; Sequence 3, Application US/09749588
; Patent No. 6423521
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: THEREOF
; FILE REFERENCE: CL001068
; CURRENT APPLICATION NUMBER: US/09/749,588
; CURRENT FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 36159
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(36159)
; OTHER INFORMATION: n = A,T,C or G
US-09-749-588-3

Query Match
Best Local Similarity 1.1%; Score 58; DB 4; Length 36159;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5450 TATTTTGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC 5507
Db 11646 TATTTTGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC 11703

Search completed: June 19, 2003, 11:55:48
Job time : 219 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:48:04 ; Search time 69.5 Seconds

(without alignments)
6581.309 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 1736

Sequence: 1 acagcagaatacatagtcca.....tcaaactcctgacctcagag 5514

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09966880/runat_14062003_175620_11056/app_query.fasta_1.5703
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=fastan -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880.scgn.1_118.ernat_14062003_175620_11056 -NCPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	0.9	585	1 ALU5_HUMAN	P39192 homo sapien
2	16	0.9	591	1 ALU8_HUMAN	P39195 homo sapien
c 3	16	0.9	591	1 ALU8_HUMAN	P39195 homo sapien
4	16	0.9	593	1 ALU7_HUMAN	P39194 homo sapien
c 5	16	0.9	593	1 ALU7_HUMAN	P39194 homo sapien
c 6	14	0.8	77	1 Q300_MOUSE	Q02722 mus musculus
c 7	14	0.8	585	1 ALU5_HUMAN	P39192 homo sapien
8	13	0.7	587	1 ALU2_HUMAN	P39189 homo sapien
c 9	13	0.7	587	1 ALU2_HUMAN	P39189 homo sapien
10	13	0.7	587	1 ALU3_HUMAN	P39190 homo sapien
c 11	13	0.7	587	1 ALU3_HUMAN	P39190 homo sapien
12	13	0.7	591	1 ALU1_HUMAN	P39188 homo sapien
c 13	13	0.7	591	1 ALU1_HUMAN	P39188 homo sapien
14	13	0.7	593	1 ALU6_HUMAN	P39193 homo sapien
c 15	13	0.7	593	1 ALU6_HUMAN	P39193 homo sapien
16	13	0.7	603	1 ALU4_HUMAN	P39191 homo sapien
c 17	13	0.7	603	1 ALU4_HUMAN	P39191 homo sapien
18	12	0.7	169	1 Y385_YEAST	P47152 saccharomyc

19	12	0.7	881	1 PKP2_HUMAN	Q99959 homo sapien
20	11	0.6	515	1 CRBA_DROME	P29747 drosophila
c 21	11	0.6	629	1 Z195_HUMAN	O14628 homo sapien
c 22	11	0.6	1411	1 TCOF_HUMAN	O13428 homo sapien
23	10	0.6	111	1 YG2C_YEAST	P53245 saccharomyc
c 24	10	0.6	167	1 YK20_YEAST	P36133 saccharomyc
c 25	10	0.6	465	1 SEP3_MOUSE	Q921S5 mus musculus
c 26	10	0.6	741	1 RED1_HUMAN	P78563 homo sapien
c 27	9	0.5	130	1 YNO3_YEAST	P53908 saccharomyc
c 28	9	0.5	139	1 YED3_YEAST	P32633 saccharomyc
c 29	9	0.5	144	1 YHP5_YEAST	P38808 saccharomyc
c 30	9	0.5	222	1 YC13_KLEPN	Q48459 klebsiella
c 31	9	0.5	231	1 NKGC_MACMU	Q9mzk6 macaca mula
c 32	9	0.5	245	1 TM4B_HUMAN	Q9ukr8 homo sapien
c 33	9	0.5	253	1 ASP_PLAFA	P13825 plasmodium
c 34	9	0.5	331	1 IF2B_MOUSE	Q99145 mus musculus
c 35	9	0.5	333	1 IF2B_HUMAN	P20042 homo sapien
c 36	9	0.5	333	1 IF2B_RABIT	P41035 cryotolagus
c 37	9	0.5	418	1 YY1_HUMAN	P49646 homo sapien
c 38	9	0.5	437	1 EF1H_XENLA	Q91375 xenopus lae
c 39	9	0.5	447	1 KBF3_HUMAN	Q04860 homo sapien
c 40	9	0.5	472	1 SX14_DROME	P40656 drosophila
c 41	9	0.5	478	1 AMP2_HUMAN	P50579 homo sapien
c 42	9	0.5	478	1 AMP2_MOUSE	O08663 mus musculus
c 43	9	0.5	478	1 AMP2_RAT	P38062 rattus norv
c 44	9	0.5	501	1 GYG2_HUMAN	O15488 homo sapien
c 45	9	0.5	529	1 NOP5_HUMAN	Q972x3 homo sapien

ALIGNMENTS

RESULT 1
ALU5_HUMAN
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentín Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS. WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U14571; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein 95 FRAME-1.
FT DOMAIN 1 95
FT DOMAIN 99 193 FRAME-2.
FT DOMAIN 197 291 FRAME-3.
FT DOMAIN 295 389 FRAME-4.
FT DOMAIN 393 487 FRAME-5.
FT DOMAIN 491 585 FRAME-6.
SQ SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.: 9.09e-08 Length: 585
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU05_HUMAN (1-585)

QY 5448 TGTATTATTAGATGGGTTTCACCATGTTGCCAGGCTGGTCT 5495
Db 446 CysilepHeSerArgAspGlyValSerProCysTrpProGlyTrpSer 461

RESULT 2

ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39155;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:836-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS. WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U14574; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.
FT DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 195 FRAME-2.
FT DOMAIN 199 294 FRAME-3.
FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Alignment Scores:
Pred. No.: 9.08e-08 Length: 591
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)
QY 5448 TGTATTTTCTAGAGATGGGTTTCACCATGTTGGCCAGGCTGTCT 5495
Db 549 CysillePheSerArgAspGlyValSerProCysrPrProGlyTrpSer 564

RESULT 3
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:103-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
```

```
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U14574; -; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 195 FRAME-2.
FT DOMAIN 199 294 FRAME-3.
FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.: 9.08e-08 Length: 591
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)
QY 5494 GACGAGCTGCGCCACATGTCGACACCCCTCTACTTAAATACAA 5447
Db 127 AspGlnProGlyGlnHisGlyClnThrProSerLeuLysIleGln 142

RESULT 4
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:103-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
```

J. Mol. Evol. 32:105-121(1991).

-1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

-1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

-1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.

-1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER, CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.

-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14573; ; NOT_ANNOTATED_CDS.

DR HYPOTHETICAL PROTEIN.

FT DOMAIN 1 97 FRAME-1.

FT DOMAIN 101 96 FRAME-2.

FT DOMAIN 200 295 FRAME-3.

FT DOMAIN 299 395 FRAME-4.

FT DOMAIN 399 494 FRAME-5.

FT DOMAIN 498 593 FRAME-6.

SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:

Pred. No.:	9.07e-08	Length:	593
Score:	16.00	Matches:	16
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.92%	Indels:	0
DB:	1	Gaps:	0

US-09-966-880A-9 (1-5514) x ALU7_HUMAN (1-593)

Qy 5448 TGTATTTTGTAGATGGTTCACCATGTGGCCAGGTGTCT 5495

Db 353 CysileheSerargaspGlyValSerProCysTrpProGlyTrpSer 368

RESULT 5

ALU7_HUMAN

ID ALU7_HUMAN STANDARD; PRT; 593 AA.

AC P39194;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Alu subfamily SQ sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-95021758; PubMed-7935834;

RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752(1994).

RN [2]

RP CONCEPT.

RX MEDLINE-92241891; PubMed-1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RL Genomics 12:838-841(1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE-88333009; PubMed-3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

RL J. Mol. Evol. 27:194-202(1988).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE-9117815; PubMed-1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.

CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER, CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.

CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14573; ; NOT_ANNOTATED_CDS.

DR HYPOTHETICAL PROTEIN.

FT DOMAIN 1 97 FRAME-1.

FT DOMAIN 101 196 FRAME-2.

FT DOMAIN 200 295 FRAME-3.

FT DOMAIN 299 395 FRAME-4.

FT DOMAIN 399 494 FRAME-5.

FT DOMAIN 498 593 FRAME-6.

SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

KW Hypothetical protein.
 FT DOMAIN 1 95
 FT DOMAIN 99 193
 FT DOMAIN 197 291
 FT DOMAIN 295 389
 FT DOMAIN 393 487
 FT DOMAIN 491 585
 SQ SEQUENCE 585 AA: 63957 MW; 468E8C4F493650A7 CRC64;
 Alignment Scores:
 Pred. No.: 1.24e-05 Length: 585
 Score: 14.00 Matches: 14
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.81% Indels: 0
 DB: 1 Gaps: 0
 US-09-966-880A-9 (1-5514) x ALU5_HUMAN (1-585)
 QY 5488 CTTGGCCAACTGCTGAACCCCTACTCTACTAAATACAA 5447
 DB 225 ProglyGlnHisGlyGluThrProSerLeuLeuLysIleGln 238
 RESULT 8
 ID ALU2_HUMAN
 AC P39189; STANDARD; PRT; 587 AA.
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Alu subfamily SB sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M.; Makalowski W.;
 RT "Alu alert.";
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 potentially misleading protein sequences.";
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation
 closely connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J.; Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -|- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 REPEATS.
 CC -|- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -|- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 ACID SEQUENCES.
 CC -|- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES.
 CC -|- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14568; -; NOT_ANNOTATED_CDS.
 KW Hypothetical protein.
 FT DOMAIN 1 96 FRAME-1.
 FT DOMAIN 100 194 FRAME-2.
 FT DOMAIN 198 292 FRAME-3.
 FT DOMAIN 296 391 FRAME-4.
 FT DOMAIN 395 489 FRAME-5.
 FT DOMAIN 493 587 FRAME-6.
 SQ SEQUENCE 587 AA: 63703 MW; 3EAB3E3E3929203 CRC64;
 Alignment Scores:
 Pred. No.: 0.000145 Length: 587
 Score: 13.00 Matches: 13
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 0.75% Indels: 0
 DB: 1 Gaps: 0
 US-09-966-880A-9 (1-5514) x ALU2_HUMAN (1-587)
 QY 3568 TGGCTCAGCGCTGTAATCCAGCACTCTGGAGGCTGAG 3606
 DB 5 TrpLeuthrProValIleProAlaLeuTrpGluAlaGlu 17
 RESULT 9
 ID ALU2_HUMAN
 AC P39189; STANDARD; PRT; 587 AA.
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Alu subfamily SB sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M.; Makalowski W.;
 RT "Alu alert.";
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 potentially misleading protein sequences.";
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation
 closely connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J.; Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -|- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 REPEATS.
 CC -|- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -|- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 ACID SEQUENCES.
 CC -|- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

RL Genomics 12:838-841(1992).

RP [3] ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation

RL closely connected with primate lineage history.";

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178815; PubMed=1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE

CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING

CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

CC REPEATS.

CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP

CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER

CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO

CC ACID SEQUENCES.

CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND

CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,

CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS

CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU

CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A

CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,

CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE

CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A

CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH

CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES

CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE

CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A

CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE

CC CODING NUCLEOTIDE SEQUENCE.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U14568; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 194 FRAME-2.

FT DOMAIN 198 292 FRAME-3.

FT DOMAIN 296 391 FRAME-4.

FT DOMAIN 395 489 FRAME-5.

FT DOMAIN 493 587 FRAME-6.

SQ SEQUENCE 587 AA; 63703 MW; 3EAAB3E3E3929203 CRC64;

Alignment Scores:

Pred. No.:	0.000145	Length:	587
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.75%	Indels:	0
DB:	1	Gaps:	0

US-09-966-880A-9 (1-5514) x ALU2_HUMAN (1-587)

Qy 3605 TCAGCCTCCAGACTGCTGGATTACAGGCTGAGCCAC 3567

|||||

Db 474 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 486

RESULT 10

ALU3_HUMAN

ID ALU3_HUMAN STANDARD; PRT; 587 AA.

AC P39190;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Alu subfamily SBI sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752(1994).

RN [2]

RP CONCEPT.

RX MEDLINE=92241891; PubMed=1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.";

RL Genomics 12:838-841(1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation

RT closely connected with primate lineage history.";

RL J. Mol. Evol. 27:194-202(1988).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178815; PubMed=1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE

CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING

CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

CC REPEATS.

CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP

CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER

CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO

CC ACID SEQUENCES.

CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND

CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,

CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS

CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU

CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A

CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,

CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE

CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A

CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH

CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES

CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE

CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A

CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE

CC CODING NUCLEOTIDE SEQUENCE.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; U14569; -: NOT_ANNOTATED_CDS.

DR Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 194 FRAME-2.

FT DOMAIN 198 292 FRAME-3.

FT DOMAIN 296 391 FRAME-4.

FT DOMAIN 395 489 FRAME-5.

FT DOMAIN 493 587 FRAME-6.

SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;

Alignment Scores:

Pred. No.:	0.000145	Length:	587
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.75%	Indels:	0
DB:	1	Gaps:	0

US-09-966-880A-9 (1-5514) x ALU3_HUMAN (1-587)

QY 3568 TGGCTCAGCCCTGTAATCCAGCACTCTGGGAGGCTGAG 3606

Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGlu 17

RESULT 11

ALU3_HUMAN

ID ALU3_HUMAN STANDARD; PRT; 587 AA.

AC P39190;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Alu subfamily S1 sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752(1994).

RN [2]

RP CONCEPT

RX MEDLINE=92241891; PubMed=1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.";

RL Genomics 12:838-841(1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation

RT closely connected with primate lineage history.";

RL J. Mol. Evol. 27:194-202(1988).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178815; PubMed=1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE

CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING

CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

CC REPEATS.

CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP

CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 CC ACID SEQUENCES.

CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS

CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,

CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; U14569; -: NOT_ANNOTATED_CDS.

DR KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 194 FRAME-2.

FT DOMAIN 198 292 FRAME-3.

FT DOMAIN 296 391 FRAME-4.

FT DOMAIN 395 489 FRAME-5.

FT DOMAIN 493 587 FRAME-6.

SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;

Alignment Scores:

Pred. No.:	0.000145	Length:	587
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.75%	Indels:	0
DB:	1	Gaps:	0

US-09-966-880A-9 (1-5514) x ALU3_HUMAN (1-587)

QY 3605 TCAGCTCCCGAGTGTGGGATTACAGCGTGAGCCAC 3567

Db 474 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 486

RESULT 12

ALU1_HUMAN

ID ALU1_HUMAN STANDARD; PRT; 591 AA.

AC P39188;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Alu subfamily J sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752(1994).

RN [2]

RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX ALU FAMILIES CLASSIFICATION.
RA MEDLINE=88333009; PubMed=3138422;
RX Quantin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
RN [1]
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14567; -; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 195 FRAME-2.
FT DOMAIN 199 294 FRAME-3.
FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;

Alignment Scores:
Pred. No.: 0.000145 Length: 591
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.75% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU1_HUMAN (1-591)
QY 3568 TGGCTCACCGCTGTATATCCAGCAGCTCTGGAGGCTGAG 3606
|||||
DB 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGlu 17

RESULT 13
ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC F39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quantin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
RN [1]
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14567; -; NOT_ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 96 FRAME-1.
CC FT DOMAIN 100 195 FRAME-2.
CC FT DOMAIN 199 294 FRAME-3.
CC FT DOMAIN 298 393 FRAME-4.
CC FT DOMAIN 397 492 FRAME-5.
CC FT DOMAIN 496 591 FRAME-6.
CC SQ SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;

Alignment Scores:
Pred. No.: 0.000145 Length: 591
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.75% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU1_HUMAN (1-591)
QY 3605 TCAGCCTCCAGAGTCGTGGGATTACAGCGGTGAGCCAC 3567
Db 378 SerAlaSerGlnSerAlaGlylleThrGlyValSerHis 390

RESULT 14
ALU6_HUMAN
ID ALU6_HUMAN STANDARD; PRT; 593 AA.
AC P39193;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SP sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178615; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
```

```
CC -----
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14572; -; NOT_ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 97 FRAME-1.
CC FT DOMAIN 101 196 FRAME-2.
CC FT DOMAIN 200 295 FRAME-3.
CC FT DOMAIN 299 395 FRAME-4.
CC FT DOMAIN 399 494 FRAME-5.
CC FT DOMAIN 498 593 FRAME-6.
CC SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.: 0.000145 Length: 593
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.75% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU6_HUMAN (1-593)
QY 3568 TGGCTCAGCGCTGTAATCCAGCACTCTGGGAGGCTGAG 3606
Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGlu 17

RESULT 15
ALU6_HUMAN
ID ALU6_HUMAN STANDARD; PRT; 593 AA.
AC P39193;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SP sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history."
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."
RL J. Mol. Evol. 32:103-121(1991).
CC -|- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -|- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -|- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -|- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -|- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14572; ; NOT_ANNOTATED_CDS.
DR DOMAIN 1 97 FRAME-1.
KW Hypothetical protein.
FT DOMAIN 101 196 FRAME-2.
FT DOMAIN 200 295 FRAME-3.
FT DOMAIN 299 395 FRAME-4.
FT DOMAIN 399 494 FRAME-5.
FT DOMAIN 498 593 FRAME-6.
SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.: 0.000145 Length: 593

Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.75% Indels: 0
DB: 1 Gaps: 0
US-09-966-880A-9 (1-5514) x ALU6_HUMAN (1-593)
QY 3605 TCAGCTCCAGAGTCTGGGATTACAGCGTGAGCCAC 3567
Db 479 SerAlaSerGlnSerAlaGlylleThrGlyValSerHis 491
Search completed: June 14, 2003, 19:19:23
Job time : 107.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:57:34 ; Search time 332 Seconds
(without alignments)
6844.240 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 1736
Sequence: 1 acagacgaatacatagtcca.....tcaaacctcgtacctcagag 5514

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues
Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_SPOOL/US09966880/runat_14062003_175621_11067/app_query.fasta_1.5703
-DB=SPTREMBL_21 -qfmt=fastan -SUFFIX=olin2p.rspt -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 -SCGN_1_1_587 -runat_14062003_175621_11067 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
c 1	22	1.3	449	4	Q9NX12	Q9nx12 homo sapien

c 2	18	1.0	101	11	Q9D443	Q9d443 mus musculu
c 3	18	1.0	104	5	Q9NLE3	Q9nle3 leishmania
c 4	18	1.0	188	5	Q9N8S9	Q9n8s9 trypanosoma
c 5	18	1.0	322	5	P90550	P90550 leishmania
c 6	18	1.0	413	12	Q9S566	Q9s566 human herpe
c 7	17	-1.0	102	4	Q8WR88	Q8wr88 homo sapien
c 8	17	1.0	112	11	Q9D324	Q9d324 mus musculu
c 9	17	1.0	535	4	Q96EB1	Q96eb1 homo sapien
c 10	16	0.9	47	4	Q96CJ4	Q96cj4 homo sapien
c 11	16	0.9	50	4	Q96EM6	Q96em6 homo sapien
c 12	16	0.9	61	4	Q9BV30	Q9bv30 homo sapien
c 13	16	0.9	73	4	Q96AN9	Q96an9 homo sapien
c 14	16	0.9	121	4	Q9HAJ0	Q9haj0 homo sapien
c 15	16	0.9	160	4	Q9H654	Q9h654 homo sapien
c 16	16	0.9	175	11	Q9D399	Q9d399 mus musculu
c 17	15	0.9	40	4	Q8WV73	Q8wv73 homo sapien
c 18	15	0.9	46	4	Q8TBW3	Q8tbw3 homo sapien
c 19	15	0.9	46	4	Q9BVC1	Q9bvc1 homo sapien
c 20	15	0.9	50	11	Q62710	Q62710 rattus norv
c 21	15	0.9	103	11	Q9CXB3	Q9cxb3 mus musculu
c 22	15	0.9	118	4	Q9H387	Q9h387 homo sapien
c 23	15	0.9	123	4	Q9HAD8	Q9had8 homo sapien
c 24	15	0.9	127	11	Q9D0V0	Q9d0v0 mus musculu
c 25	15	0.9	133	4	Q96JR5	Q96jrs homo sapien
c 26	15	0.9	143	11	Q64150	Q64150 mus musculu
c 27	15	0.9	257	4	Q9H7F8	Q9h7f8 homo sapien
c 28	14	0.8	50	11	Q8R193	Q8r193 mus musculu
c 29	14	0.8	72	4	Q9UI79	Q9ui79 homo sapien
c 30	14	0.8	128	11	Q9LWM8	Q9lwm8 mus musculu
c 31	14	0.8	139	4	Q9GZ25	Q9gz25 homo sapien
c 32	14	0.8	207	11	Q9D3D2	Q9d3d2 mus musculu
c 33	14	0.8	231	4	Q9NR08	Q9nr08 homo sapien
c 34	14	0.8	318	4	Q96QL7	Q96ql7 homo sapien
c 35	14	0.8	375	4	Q60448	Q60448 homo sapien
c 36	14	0.8	423	4	Q9NWF0	Q9nwf0 homo sapien
c 37	14	0.8	461	5	Q9NWX3	Q9nwx3 leishmania
c 38	13	0.7	29	4	Q96IR5	Q96ir5 homo sapien
c 39	13	0.7	90	4	Q96IG1	Q96ig1 homo sapien
c 40	13	0.7	130	4	Q9HBS7	Q9hbs7 homo sapien
c 41	13	0.7	174	4	Q9H926	Q9h926 homo sapien
c 42	13	0.7	208	4	Q9GQ97	Q9gq97 homo sapien
c 43	13	0.7	232	4	Q9H5R3	Q9h5r3 homo sapien
c 44	13	0.7	296	4	Q8WWD5	Q8wwd5 homo sapien
c 45	13	0.7	296	7	Q29890	Q29890 homo sapien

ALIGNMENTS

RESULT 1

Q9NX12	PRELIMINARY;	PRT;	449 AA.
ID	Q9NX12		
AC	Q9NX12:		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	CDNA FLJ20234 fis, clone COLF5673.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=COLON MUCOSA;		
RA	Watanabe K., Kumeagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,		
RA	Suzuki Y., Ohtsubashi M., Nishi T., Shibahara T., Tanaka T.,		
RA	Nakamura Y., Isogai T., Sugano S.;		
RT	"NEO human cDNA sequencing project."		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK000241; BAA91028.1;		
DR	InterPro; IPR001849; PH.		
DR	InterPro; IPR000980; SH2.		
DR	Pfam; PF00017; SH2; 1.		
DR	SMART; SM00233; PH; 1.		

[illegible]

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:09:34 ; Search time 52.5 Seconds
(without alignments)
6180.495 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 1736

Sequence: 1 acagacgaatacatagctgtccca.....tcaaaactcctgacctcagag 5514

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_Spool/US09966880/runat_14062003_175623_11106/app_query.fasta_1.5703
-DB=Issued_Patents_AA -QMT=fastan -SUFFIX=olin2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 -CGN_1_1_83@runat_14062003_175623_11106 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	14	0.8	96	4	US-09-227-357-171
2	14	0.8	95	4	US-09-605-785-827
3	14	0.8	224	4	US-09-605-785-825
c 4	14	0.8	375	2	US-08-454-557C-121
c 5	14	0.8	375	2	US-08-340-426D-121
c 6	14	0.8	375	2	US-08-450-673C-121
c 7	14	0.8	410	1	US-08-153-848-7
c 8	14	0.8	410	3	US-09-299-843A-7
c 9	14	0.8	410	4	US-09-088-337B-7
c 10	14	0.8	410	5	PCT-US93-11153-7
c 11	14	0.8	1079	3	US-09-058-489-22
12	14	0.8	1261	4	US-09-605-785-538

13	14	0.8	1261	4	US-09-439-313-538	Sequence 538, App
14	13	0.7	65	4	US-09-227-357-538	Sequence 538, App
c 15	13	0.7	70	4	US-09-188-930-131	Sequence 131, App
16	12	0.7	90	4	US-09-227-357-171	Sequence 171, App
17	12	0.7	122	4	US-09-227-357-285	Sequence 285, App
18	12	0.7	397	5	PCT-US95-17111A-121	Sequence 121, App
c 19	12	0.7	605	4	US-09-440-936-2	Sequence 2, Appli
20	11	0.6	52	4	US-09-227-357-537	Sequence 537, App
21	11	0.6	521	2	US-08-721-684C-2	Sequence 2, Appli
22	11	0.6	521	2	US-09-005-970-2	Sequence 2, Appli
23	11	0.6	521	4	US-09-407-715-2	Sequence 2, Appli
c 24	11	0.6	1079	3	US-09-058-489-22	Sequence 22, Appli
c 25	10	0.6	19	4	US-09-463-238-30	Sequence 30, Appli
c 26	10	0.6	32	4	US-09-288-143-171	Sequence 171, App
27	10	0.6	35	4	US-09-288-143-170	Sequence 170, App
c 28	10	0.6	36	1	US-07-948-357-12	Sequence 12, Appli
c 29	10	0.6	36	1	US-07-948-357-13	Sequence 13, Appli
c 30	10	0.6	36	2	US-08-450-417-12	Sequence 12, Appli
c 31	10	0.6	36	2	US-08-450-417-13	Sequence 13, Appli
c 32	10	0.6	36	3	US-08-449-741-12	Sequence 12, Appli
c 33	10	0.6	36	3	US-08-449-741-13	Sequence 13, Appli
c 34	10	0.6	36	3	US-08-449-754-12	Sequence 12, Appli
c 35	10	0.6	36	3	US-08-449-754-13	Sequence 13, Appli
c 36	10	0.6	36	4	US-09-465-646-12	Sequence 12, Appli
c 37	10	0.6	36	4	US-09-465-646-13	Sequence 13, Appli
38	10	0.6	49	4	US-09-288-143-93	Sequence 93, Appli
39	10	0.6	56	4	US-09-227-357-577	Sequence 577, App
40	10	0.6	80	4	US-09-227-357-249	Sequence 249, App
c 41	10	0.6	99	4	US-09-288-143-168	Sequence 168, App
c 42	10	0.6	99	4	US-09-288-143-168	Sequence 168, App
c 43	10	0.6	182	1	US-08-226-264-28	Sequence 28, Appli
c 44	10	0.6	312	4	US-09-254-465A-9	Sequence 9, Appli
45	10	0.6	388	4	US-09-265-630-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1

US-09-227-357-171
; Sequence 171, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; EARLIER FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08

```

; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals stop translation
US-09-227-357-171

```

```

Alignment Scores:
Pred. No.: 0.000128 Length: 90
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

```

US-09-966-880a-9 (1-5514) x US-09-227-357-171 (1-90)

```

QY 3605 TCAGCCTCCAGAGTCGTGGATTACAGCGTGACCCAC 3564
|||||
DB 36 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 49
|||||

```

RESULT 2

```

; Sequence 827, Application us/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.

```

```

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 827
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-827

Alignment Scores:
Pred. No.: 0.000127 Length: 96
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880a-9 (1-5514) x US-09-605-785-827 (1-96)

QY 5462 AGATGGGTTTCACCATGTGGCCAGCGTGTCAAACTCC 5503
|||||
DB 75 ArgtrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 88
|||||

RESULT 3
US-09-605-785-825
; Sequence 825, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 825
; LENGTH: 224

```

;
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-825

Alignment Scores:
Pred. No.: 0.000118 Length: 224
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-605-785-825 (1-224)

QY 5462 AGATGGGTTTCACCATGTTGCGAGGTGCTCAAACTCC 5503
|||||
Db 211 ArgTrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 224
|||||

RESULT 4
US-08-454-557C-121
; Sequence 121, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,557C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3840003
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 121:
SEQUENCE CHARACTERISTICS:
LENGTH: 375 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-454-557C-121

Alignment Scores:
Pred. No.: 0.000112 Length: 375
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-454-557C-121 (1-375)

QY 3605 TCAGCCTCCAGAGTCTGGGATTACAGCGGTGAGCCACCAC 3564
|||||
Db 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 292
|||||

US-08-454-557C-121
; Sequence 121, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:

RESULT 5
US-08-340-426D-121
; Sequence 121, Application US/08340426D
; Patent No. 5948634
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,426D
FILING DATE: 14-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3840002
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 121:
SEQUENCE CHARACTERISTICS:
LENGTH: 375 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-340-426D-121

Alignment Scores:
Pred. No.: 0.000112 Length: 375
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-340-426D-121 (1-375)

QY 3605 TCAGCCTCCAGAGTCTGGGATTACAGCGGTGAGCCACCAC 3564
|||||
Db 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 292
|||||

US-08-450-673C-121
; Sequence 121, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:

RESULT 6
US-08-450-673C-121
; Sequence 121, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-673C-121

Alignment Scores:
Pred. No.: 0.000112 Length: 375
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-450-673C-121 (1-375)

QY 3605 TCAGCTCCAGAGTGGGATTACAGCGGTGAGCCAC 3564
|||||
Db 279 SerAlaSerGlnSerIaGlylleThrGlyValSerHisHis 292

RESULT 7
US-08-153-848-7
; Sequence 7, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5759804and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
```

```
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-153-848-7

Alignment Scores:
Pred. No.: 0.000111 Length: 410
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-153-848-7 (1-410)

QY 3291 ATGATTGCACCTGCTCCAGCTGGGTGAGCGGTGAGA 3250
|||||
Db 1 MetIleAlaProLeuHisSerSerLeuGlyAsnArgValArg 14

RESULT 8
US-09-299-843A-7
; Sequence 7, Application US/09299843A
; Patent No. 6107475
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,843A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 09/088,337
; FILING DATE: 01-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,848
; FILING DATE: 17-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Jill E. Uhl
; REGISTRATION NUMBER: 43,213
; REFERENCE/DOCKET NUMBER: 27866/32059B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

us-09-966-880a-9.olin2p.ra1

Fri Jun 20 09:21:58 2003

MOLECULE TYPE: protein
US-09-299-843A-7

Alignment Scores: 0.000111 Length: 410
Pred. No.: 14.00 Matches: 14
Score: 100.00% Conservativity: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 0.81% Gaps: 0
DB:

US-09-966-880A-9 (1-5514) x US-09-299-843A-7 (1-410)

QY 3291 ATGATTGCACCTGCCTCCAGCTGGTAAACAGGGTGAGA 3250

Db 1 MetileAlaProleuHisSerLeuGlyAsnArgValArg 14

RESULT 9

US-09-088-337B-7

; Sequence 7, Application US/09088337B

; Patent No. 6348574

; GENERAL INFORMATION:

; APPLICANT: Godiska, Ronald

; Gray, Patrick W.

; Schweikart, Vicki L.

; TITLE OF INVENTION: No. 6348574el Seven Transmembrane Receptors

; NUMBER OF SEQUENCES: 66

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; Borun

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/088.337B

; FILING DATE: 01-Jun-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/153,848

; FILING DATE: 17-NOV-1993

; APPLICATION NUMBER: US 07/977,452

; FILING DATE: 17-NOV-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6348574and, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 31794

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (312) 474-6300

; TELEFAX: (312) 474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 410 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-088-337B-7

Alignment Scores: 0.000111 Length: 410
Pred. No.: 14.00 Matches: 14
Score: 100.00% Conservativity: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 0.81% Gaps: 0
DB:

US-09-966-880A-9 (1-5514) x US-09-088-337B-7 (1-410)

QY 3291 ATGATTGCACCTGCCTCCAGCTGGTAAACAGGGTGAGA 3250

Db 1 MetileAlaProleuHisSerLeuGlyAsnArgValArg 14

RESULT 10

PCT-US93-11153-7

; Sequence 7, Application PC/TUS9311153

; GENERAL INFORMATION:

; APPLICANT: Godiska, Ronald

; Gray, Patrick W.

; APPLICANT: Schweikart, Vicki L.

; TITLE OF INVENTION: Novel Seven Transmembrane Receptors

; NUMBER OF SEQUENCES: 64

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Bicknell

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/11153

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/977,452

; FILING DATE: 17-NOV-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Noland, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 31794

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (312) 474-6300

; TELEFAX: (312) 474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 410 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US93-11153-7

Alignment Scores: 0.000111 Length: 410
Pred. No.: 14.00 Matches: 14
Score: 100.00% Conservativity: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 0.81% Gaps: 0
DB:

US-09-966-880A-9 (1-5514) x PCT-US93-11153-7 (1-410)

QY 3291 ATGATTGCACCTGCCTCCAGCTGGTAAACAGGGTGAGA 3250

Db 1 MetileAlaProleuHisSerLeuGlyAsnArgValArg 14

RESULT 11

US-09-059-489-22

; Sequence 22, Application US/09058489

; Patent No. 6103886

; GENERAL INFORMATION:

; APPLICANT: Whitehead Institute for Biomedical Research

; APPLICANT: Lahn, Bruce

; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: the Y Chromosome
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:
Pred. No.: 0.000102 Length: 1079
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 3 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-058-489-22 (1-1079)
QY 3605 TCAGCTCCAGAGTCTGGATTACAGCGGTGAGCCAC 3564
Db 1064 SerAtaSerGlnSerAlaGlyIleThrGlyValSerHis 1077

RESULT 12
US-09-605-785-538
; Sequence 538, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 538
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-538

Alignment Scores:
Pred. No.: 0.0001 Length: 1261
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-605-785-538 (1-1261)
QY 5462 AGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAGTCC 5503
Db 1248 ArgTrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 1261
RESULT 13
US-09-439-313-538
; Sequence 538, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 538
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-313-538

Alignment Scores:
Pred. No.: 0.0001 Length: 1261
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-439-313-538 (1-1261)
QY 5462 AGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAGTCC 5503
Db 1248 ArgTrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 1261

RESULT 14
US-09-227-357-538
; Sequence 538, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08

	EARLIER APPLICATION NUMBER:	60/051,931	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,932	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,916	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,930	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,918	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,920	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/052,733	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/052,795	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,919	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/051,928	
	EARLIER FILING DATE:	1997-07-08	
	EARLIER APPLICATION NUMBER:	60/055,722	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,723	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,948	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,949	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,953	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,950	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,947	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,964	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/056,360	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,684	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,984	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/055,954	
	EARLIER FILING DATE:	1997-08-18	
	EARLIER APPLICATION NUMBER:	60/058,785	
	EARLIER FILING DATE:	1997-09-12	
	EARLIER APPLICATION NUMBER:	60/058,664	
	EARLIER FILING DATE:	1997-09-12	
	EARLIER APPLICATION NUMBER:	60/058,660	
	EARLIER FILING DATE:	1997-09-12	
	EARLIER APPLICATION NUMBER:	60/058,661	
	EARLIER FILING DATE:	1997-09-12	
	NUMBER OF SEQ ID NOS:	672	
	SOFTWARE:	PatentIn Ver. 2.0	
	SEQ ID NO 538		
	LENGTH:	65	
	TYPE:	PRT	
	ORGANISM:	Homo sapiens	
	US-09-227-357-538		
	Alignment Scores:	0.00118	65
	Pred. No.:	13.00	13
	Score:	Conservative:	0
	Percent Similarity:	100.00%	0
	Best Local Similarity:	100.00%	0
	Query Match:	0.75%	0
	Gaps:	4	0
	DB:		
	US-09-966-880A-9 (1-5514) x US-09-227-357-538 (1-65)		
QY	357TCTATCCAGCACTCTCGGAGGTGGTGGA	361S	
Db	43 ProValIleProAlaLeuTpGluAgluValgIcgly	55	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 20:16:19 ; Search time 576.918 seconds
(without alignments)
11000.051 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 2818
Sequence: 1 agagaacatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	2818	100.0	2818	21	AAC55312	Human activation-i
2	2174.6	77.2	6564	21	AAC55314	Human activation-i
3	2174.6	77.2	11204	21	AAC55339	Human activation-i
4	2172	77.1	2172	21	AAC55313	Human activation-i
5	1489.2	52.8	1665	22	AAK81088	Human immune/haema
6	603.4	21.4	2440	21	AAC55307	Mouse activation-i
7	574	20.4	574	22	AAK81089	Human immune/haema
8	429	15.2	429	22	AAK61819	Human immune/haema
9	295.2	10.5	13862	22	ABA08208	Human ovarian and

10	295.2	10.5	13862	22	AAI02789	Human reproductive
11	295.2	10.5	13862	22	AAI07516	Human reproductive
12	292.6	10.4	7809	22	ABA15724	Human nervous syst
13	292.6	10.4	7809	22	AAI03406	Human reproductive
14	292.4	10.4	30393	22	AAK67239	Human immune/haema
15	291	10.3	25012	22	ABA15431	Human nervous syst
16	286.2	10.2	37314	22	AAK71358	Human immune/haema
17	286.2	10.2	172570	24	AAQ88207	Human osteoblast d
18	284.4	10.1	32249	22	AAI04676	Human reproductive
19	284.4	10.1	32249	23	ABL97583	Human testicular a
20	284.4	10.1	129722	24	ABQ88117	Human osteoblast d
21	283.4	10.1	1792	24	ABA96614	Human alpha interf
22	283.4	10.0	22428	22	AAI1759	Genomic sequence #
23	282.4	10.0	13919	24	ABK86218	DNA encoding AIP-1
24	282.4	10.0	13919	24	ABK86220	AIP-1/FLASH promot
25	282.4	10.0	13919	24	ABK86221	AIP-1/FLASH promot
26	282.4	10.0	21404	24	ABK86229	AIP-1/FLASH promot
27	281.4	10.0	65608	24	ABL62910	Breast cancer rela
28	281.4	10.0	65608	24	ABL64414	Stomach cancer rel
29	281.4	10.0	65608	24	ABL67668	Oesophagus cancer
30	281	10.0	23452	22	AAI42122	Genomic sequence #
31	280.2	9.9	11821	22	AAI36492	Human musculoskele
32	280.2	9.9	122888	24	ABK83569	Human CDNA differe
33	280	9.9	15297	24	ABA94501	Human proto-oncoge
34	279.6	9.9	23457	22	AAI42121	Genomic sequence #
35	279.6	9.9	23458	22	AAI42120	Genomic sequence #
36	278.4	9.9	12263	24	ABK84514	Human CDNA differe
37	278.4	9.9	21596	22	AAK71582	Human immune/haema
38	278.2	9.9	32204	22	AAI05849	Human reproductive
39	278.2	9.9	32204	23	ABL98413	Human testicular a
40	277.6	9.9	86080	24	ABQ88164	Human osteoblast d
41	277.6	9.9	86080	24	ABK83561	Human CDNA differe
42	277.2	9.8	5253	22	AAK89859	Human digestive sy
43	277.2	9.8	5257	22	AAK89860	Human digestive sy
44	277	9.8	12221	22	AAK71585	Human immune/haema
45	277	9.8	30393	22	AAK67239	Human immune/haema

ALIGNMENTS

RESULT 1
AAC55312
ID AAC55312 standard; cDNA; 2818 BP.
XX AC AAC55312;
XX DT 05-FEB-2001 (first entry)
XX DE Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.
XX DE Human activation-induced cytidine deaminase; AID: cytidine deaminase;
KW Immune related disease; allergy; allergic disease; antiallergic;
KW antianemic; antitubercular; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.
XX OS Homo sapiens.

Key Location/Qualifiers
CDS 80..676
/*tag= a
/product= "activation-induced cytidine deaminase"

WO2000058480-A1.

05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP01918.
 XX 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX (NISB) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX Honjo T, Muramatsu M;
 PI WPI: 2000-611715/58.
 XX P-PSDB; AAB24198.
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as
 DR a target for drug development for immune-related diseases including
 XX allergies -
 PT Claim 3; Page 135-139; 174pp; Japanese.
 PS The present sequence encodes human activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,
 CC antinaemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, Diegeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;
 SQ Query Match 100.0%; Score 2818; DB 21; Length 2818;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGAGAACCATTAATTAAGTGAGATTTTCTGGCCGTGAGACTTGCAGGGAGGCAAGA 60
 DB 1 AGAGAACCATTAATTAAGTGAGATTTTCTGGCCGTGAGACTTGCAGGGAGGCAAGA 60
 QY 61 AGACACTCTGGACACCATATGACAGCCCTTGTATGAACCGGAGGAAGTTCTTTACCA 120
 DB 61 AGACACTCTGGACACCATATGACAGCCCTTGTATGAACCGGAGGAAGTTCTTTACCA 120
 QY 121 ATTCAAAATATGTCGGCTGAGCTAAGGCTGGCGTGAGACTACCTGTGTACGTAGTAA 180
 DB 121 ATTCAAAATATGTCGGCTGAGCTAAGGCTGGCGTGAGACTACCTGTGTACGTAGTAA 180
 QY 181 GAGCGTGACAGTGCTACATCTCTTTTCACTGGACTTTGGTATCTTGGCAATGAACGG 240
 DB 181 GAGCGTGACAGTGCTACATCTCTTTTCACTGGACTTTGGTATCTTGGCAATGAACGG 240
 QY 241 CTGCCACGTGAATTCCTCTCTCCGCTACATCTCGGACTGGGACTAGACCCCTGGCCG 300
 DB 241 CTGCCACGTGAATTCCTCTCTCCGCTACATCTCGGACTGGGACTAGACCCCTGGCCG 300
 QY 301 CTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTGTGTACGACTGTGCCCGACATGT 360
 DB 301 CTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTGTGTACGACTGTGCCCGACATGT 360
 QY 361 GGCCGACTTTCTCGGAGGAAACCCCAACCTCAGTCTGAGATCTTCCACCGCGCCCTCTA 420
 DB 361 GGCCGACTTTCTCGGAGGAAACCCCAACCTCAGTCTGAGATCTTCCACCGCGCCCTCTA 420
 QY 421 CTTCTGTGAGACCGAAGGCTCAGCCCGGAGGGCTCGCGGCTGCACCGCCCGGGGT 480
 DB 421 CTTCTGTGAGACCGAAGGCTCAGCCCGGAGGGCTCGCGGCTGCACCGCCCGGGGT 480

DB 421 CTTCTGTGAGACCGAAGGCTCAGCCCGGAGGGCTCGCGGCTGCACCGCCCGGGGT 480
 QY 481 GCAAAATAGCATCATGACCTTCAAGATATATTTTACTGCTGCTGAATACATTTTCTAGAAAA 540
 DB 481 GCAAAATAGCATCATGACCTTCAAGATATATTTTACTGCTGCTGAATACATTTTCTAGAAAA 540
 QY 541 CCATGAAAGAACTTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCTCCAG 600
 DB 541 CCATGAAAGAACTTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCTCCAG 600
 QY 601 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACCATTTTCG 660
 DB 601 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACCATTTTCG 660
 QY 661 TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACATGAAATATCTCTGCTG 720
 DB 661 TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACATGAAATATCTCTGCTG 720
 QY 721 AAGACAGTGGATAAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCAC 780
 DB 721 AAGACAGTGGATAAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCAC 780
 QY 781 TTTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGCTTG 840
 DB 781 TTTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGCTTG 840
 QY 841 AAAATAGAGAAGAACACACAGGCTCTGCCAGGAGCTGCTGCAATTTGGTGCAGTTTGAAT 900
 DB 841 AAAATAGAGAAGAACACACAGGCTCTGCCAGGAGCTGCTGCAATTTGGTGCAGTTTGAAT 900
 QY 901 GCAACATTTGCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGT 960
 DB 901 GCAACATTTGCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGT 960
 QY 961 CAACGTTTTCTATGACTTTTAGTGGATGAGAGCAGAGGTAGATCTTAAAAAGCATG 1020
 DB 961 CAACGTTTTCTATGACTTTTAGTGGATGAGAGCAGAGGTAGATCTTAAAAAGCATG 1020
 QY 1021 GTGAGAGGATCAAAATGTTTTTATCAACATCTTTTATTTATTTGATTCAATTTGAGTTAAC 1080
 DB 1021 GTGAGAGGATCAAAATGTTTTTATCAACATCTTTTATTTATTTGATTCAATTTGAGTTAAC 1080
 QY 1081 AGTGGTTAGTGATAGATTTTCTATCTTTTCCCTTGACGTTTACTTTCAAGTAAAC 1140
 DB 1081 AGTGGTTAGTGATAGATTTTCTATCTTTTCCCTTGACGTTTACTTTCAAGTAAAC 1140
 QY 1141 AAACCTCTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGTATCTGGGTGATTGT 1200
 DB 1141 AAACCTCTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGTATCTGGGTGATTGT 1200
 QY 1201 GACCCCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCTGTATGTTTAAATCAG 1260
 DB 1201 GACCCCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCTGTATGTTTAAATCAG 1260
 QY 1261 CAGAACATGTTTTTATGTTTGTACAAAAGAGATTGTTATGGGTGGGATGAGGATATA 1320
 DB 1261 CAGAACATGTTTTTATGTTTGTACAAAAGAGATTGTTATGGGTGGGATGAGGATATA 1320
 QY 1321 GACCATGCTGGTGCACCTTCAAGTCTTAAATAAAGAGTCTTAAATAAGGAGGAGGAG 1380
 DB 1321 GACCATGCTGGTGCACCTTCAAGTCTTAAATAAAGAGTCTTAAATAAGGAGGAGGAG 1380
 QY 1381 TGTGAACAAGACACCCCTTAATATGAGTCTGAAAGTAGCAAAATCTTCTGGAACGC 1440
 DB 1381 TGTGAACAAGACACCCCTTAATATGAGTCTGAAAGTAGCAAAATCTTCTGGAACGC 1440
 QY 1441 AAACCTCTTTTAAAGAGTCCCTAAATTTAGAACACCCCAAACTTCCACATATCAATTA 1500
 DB 1441 AAACCTCTTTTAAAGAGTCCCTAAATTTAGAACACCCCAAACTTCCACATATCAATTA 1500
 QY 1501 GCAAAACAATTTGGAAGGAGTTGCTTGAATGTTGGGAGAGGAGGAGGAGGAGGAGGAG 1560
 DB 1501 GCAAAACAATTTGGAAGGAGTTGCTTGAATGTTGGGAGAGGAGGAGGAGGAGGAGGAG 1560

QY	1561	GGGCTCTCTCATCTCAGAAATGCCAATCAAGTTCAGGTTCGTACATAATTTTGTAATGTGTGT	1620
DB	1561	GGGCTCTCTCATCTCAGAAATGC CAATCAAGTTCAGGTTCGTACATAATTTTGTAATGTGTGT	1620
QY	1621	GATGCTTCTCCCAAAGGTATATTAACTATATATAAGAGAGTTGTGCACAAAACAGAAATCATAA	1680
DB	1621	GATGCTTCTCCCAAAGGTATATTAACTATATATAAGAGAGTTGTGCACAAAACAGAAATCATAA	1680
QY	1681	AGCTGCGAACCGTGGCACACGCTCATAGTTCTTAGCTGCTTGGGAGGTTGAGGAGGAGGA	1740
DB	1681	AGCTGCGAACCGTGGCACACGCTCATAGTTCTTAGCTGCTTGGGAGGTTGAGGAGGAGGA	1740
QY	1741	TGGCTTTGAACACAGGTGTTCAAGCCACGCTGGGCAACATACAAGAATCCTGTCCTCTCAA	1800
DB	1741	TGGCTTTGAACACAGGTGTTCAAGCCACGCTGGGCAACATACAAGAATCCTGTCCTCTCAA	1800
QY	1801	AAAAAAAAAAAAAAGAACAGAGAGGGCGGGCTGGTGGCTGCAGCCCTGTAATCC	1860
DB	1801	AAAAAAAAAAAAAAGAACAGAGAGGGCGGGCTGGTGGCTGCAGCCCTGTAATCC	1860
QY	1861	CAGCACTTTGGGAGCGGAGCGGGCGGATCACTGTGGTCAAGAGTTTTGAGACCAGCCT	1920
DB	1861	CAGCACTTTGGGAGCGGAGCGGGCGGATCACTGTGGTCAAGAGTTTTGAGACCAGCCT	1920
QY	1921	GGCCAACATGGCAAAACCCGCTCTGTACTCAAATGC AAAATAGCCAGGCGTGTAGC	1980
DB	1921	GGCCAACATGGCAAAACCCGCTCTGTACTCAAATGC AAAATAGCCAGGCGTGTAGC	1980
QY	1981	AGGCACCTGTAATCCCACGCTACTTGGGAGGCTCAGGCAGGAGAAATCGCTTGAACCCAGGA	2040
DB	1981	AGGCACCTGTAATCCCACGCTACTTGGGAGGCTCAGGCAGGAGAAATCGCTTGAACCCAGGA	2040
QY	2041	GGTGGAGGTTGCAGTAACTGAGATCGTGCCGTTGCATCCAGCCTGGGGGACAGAGCA	2100
DB	2041	GGTGGAGGTTGCAGTAACTGAGATCGTGCCGTTGCATCCAGCCTGGGGGACAGAGCA	2100
QY	2101	AGACTCTGCTCAGAAAAA AAAAAAGACAGAGAGAGAAAAAGAGAACAAATATTTC	2160
DB	2101	AGACTCTGCTCAGAAAAA AAAAAAGACAGAGAGAGAAAAAGAGAACAAATATTTC	2160
QY	2161	GGAGAAGGATGGGAAGCATTCGAAGGAAATGTGCTTTATCCAAACAAATGTAAGGA	2220
DB	2161	GGAGAAGGATGGGAAGCATTCGAAGGAAATGTGCTTTATCCAAACAAATGTAAGGA	2220
QY	2221	GCCAAATAGGGATCCCTATTGTCTCTTTGGTGCTATTTGTGCCCTAACAACTGCTTTT	2280
DB	2221	GCCAAATAGGGATCCCTATTGTCTCTTTGGTGCTATTTGTGCCCTAACAACTGCTTTT	2280
QY	2281	GACAGTGAGAAAAATATTCAGAAATACCATATCCCTGTGCCGTTATTACCTACCAACCCT	2340
DB	2281	GACAGTGAGAAAAATATTCAGAAATACCATATCCCTGTGCCGTTATTACCTACCAACCCT	2340
QY	2341	TGCAATGAAGATGAGCAGATCCACAGSAAACTTGAATGCACAACTGCTCTATTATTTAATC	2400
DB	2341	TGCAATGAAGATGAGCAGATCCACAGSAAACTTGAATGCACAACTGCTCTATTATTTAATC	2400
QY	2401	TTATTGTACATAAGTTTGTAAAAAGTTAAAAATTTGTTACTTCATGTATTCATTTATATT	2460
DB	2401	TTATTGTACATAAGTTTGTAAAAAGTTAAAAATTTGTTACTTCATGTATTCATTTATATT	2460
QY	2461	TTATATTATTTTCGGCTAATGATTTTTTATTAACATGATTTCTCTTTCTCGATATATTGA	2520
DB	2461	TTATATTATTTTCGGCTAATGATTTTTTATTAACATGATTTCTCTTTCTCGATATATTGA	2520
QY	2521	AATGGAGTCTCAAAGCTTCTAAATTTATAACTTTTAGAAATGATTTCTTAATAACAACGAT	2580
DB	2521	AATGGAGTCTCAAAGCTTCTAAATTTATAACTTTTAGAAATGATTTCTTAATAACAACGAT	2580
QY	2581	GTAAITGTACATTTGCAGTAATCGTGCTACGAAGCCATTTCTCTGATTTTTAGTAAACT	2640
DB	2581	GTAAITGTACATTTGCAGTAATCGTGCTACGAAGCCATTTCTCTGATTTTTAGTAAACT	2640

Qy	2641	TTTATGACAGCAAAATTTGGCTCTGGCTCAGCTTTCAATCAGTTAAATAAATGATAATAAT	2760
Db	2641	TTTATGACAGCAAAATTTGGCTCTGGCTCAGCTTTCAATCAGTTAAATAAATGATAATAAT	2700
Qy	2701	TTTGGAGCGTGTGAAGATAAAATACCAAAATAAAATAATATAAAAAGTCATTATATGAAGT	2760
Db	2701	TTTGGAGCGTGTGAAGATAAAATACCAAAATAAAATAATATAAAAAGTCATTATATGAAGT	2760
Qy	2761	TAAATAAAAAATCAGTATGATGAATAAACTTTGAAAAAATAAAAAAATAAAAAA	2818
Db	2761	TAAATAAAAAATCAGTATGATGAATAAACTTTGAAAAAATAAAAAAATAAAAAA	2818
RESULT 2			
AAC55314			
ID	AAC55314	standard; DNA: 6564 BP.	
XX			
AC	AAC55314;		
XX			
DT	05-FEB-2001	(first entry)	
XX			
DE	Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.		
XX			
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;		
KW	immune related disease; allergy; allergic disease; anti-allergic;		
KW	antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;		
KW	gene therapy; B cell associated immune system disorder; food allergy;		
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;		
KW	IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;		
KW	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;		
KW	ataxia telangiectasia; common variable immunodeficiency disorder;		
KW	major histocompatibility class II deficiency disease;		
KW	auto immunodeficiency syndrome; IgG subclass selection disorder; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200058480-A1.		
XX			
PD	05-OCT-2000.		
XX			
PF	28-MAR-2000; 2000WO-JP01918.		
XX			
PR	29-MAR-1999; 99JP-0087192.		
PR	24-JUN-1999; 99JP-0178999.		
PR	27-DEC-1999; 99JP-0371382.		
XX			
PA	(NLSB) JAPAN TOBACCO INC.		
PA	(HONJ/) HONJO T.		
XX			
PI	Honjo T, Muramatsu M;		
XX			
DR	WPI; 2000-611715/58.		
XX			
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as		
PT	a target for drug development for immune-related diseases including		
PT	allergies.-		
PS			
PS	Claim 17; Page 145-150; 174pp; Japanese.		
XX			
CC	The present invention describes an activation-induced cytidine deaminase		
CC	(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and		
CC	has cytidine activity similar to APOBEC-1. AID has anti-allergic,		
CC	antianemic, antiasthmatic, ophthalmological, anti-HIV and		
CC	dermatological activities, and can be used in gene therapy. AID		
CC	polynucleotides are useful in methods for identifying drugs for the		
CC	treatment of B cell associated immune system disorders, immunodeficiency		
CC	diseases and allergies, such as immunoglobulin A (IgA) deficiency		
CC	disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic		
CC	colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen		
CC	disease, DiGeorge disease, ataxia telangiectasia, common variable		
CC	immunodeficiency disorder, MHC (major histocompatibility class) class		
CC	II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated		
CC	IgE disorder, and IgG subclass selection disorder. The DNA sequences		

CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents a genomic DNA sequence of human AID.
 XX
 SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 other;

Query Match 77.2%; Score 2174.6; DB 21; Length 6564;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	617	CTTTTCCCTCTGATGAGTTGATGACTTACGAGACGCAATTCGTTGGGACTTTGA	676
DB	3735	CTACGCCCTGATGAGTTGATGACTTACGAGACGCAATTCGTTGGGACTTTGA	3794
QY	677	TACCAACTTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGACAGTGGATAAAA	736
DB	3795	TACCAACTTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGACAGTGGATAAAA	3854
QY	737	ACAGTCCCTCAAGTCTTCTCTGTTTTTATCTTCAACTCTCATTCTTCTTAGAGTTTACA	796
DB	3855	AACAGTCCCTCAAGTCTTCTCTGTTTTTATCTTCAACTCTCATTCTTCTTAGAGTTTACA	3914
QY	797	GAATAAATATTTATATACGACTCTTTAAAGAGATCTATGCTTGAATAAGAGAAC	856
DB	3915	GAATAAATATTTATACGACTCTTTAAAGAGATCTATGCTTGAATAAGAGAAC	3974
QY	857	ACAGTCTGGCCAGGACGCTGTCGAATTTGGTGCAGTTTGAATGCAACATTCGCCCTA	916
DB	3975	ACAGTCTGGCCAGGACGCTGTCGAATTTGGTGCAGTTTGAATGCAACATTCGCCCTA	4034
QY	917	CTGGGAATACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCTATGA	976
DB	4035	CTGGGAATACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCTATGA	4094
QY	977	CTTTTAGTAGGATGAGACAGAGTATGATCTTAAAGAGATCTTAAAGTGTCAACGTTTTCTATGA	1036
DB	4095	CTTTTAGTAGGATGAGACAGAGTATGATCTTAAAGAGATCTTAAAGTGTCAACGTTTTCTATGA	4154
QY	1037	TTTTTATATCAACATCCTTTATTTATTTGATTTCAATTTGAGTTAAACAGTGGTTAGTGATA	1096
DB	4155	TTTTTATATCAACATCCTTTATTTATTTGATTTCAATTTGAGTTAAACAGTGGTTAGTGATA	4214
QY	1097	GATTTTCTATTTCTTTCCCTTGACGTTTACTTTTCAAGTAACACAACTCTTCCATCAGG	1156
DB	4215	GATTTTCTATTTCTTTCCCTTGACGTTTACTTTTCAAGTAACACAACTCTTCCATCAGG	4274
QY	1157	CCATGATCTATAGGACCTCTTAATGAGATGATCTGGGTGATTTGACCCCAACCATCTC	1216
DB	4275	CCATGATCTATAGGACCTCTTAATGAGATGATCTGGGTGATTTGACCCCAACCATCTC	4334
QY	1217	TCCAAAGCATTAATATCCCAATCATGCGCTGATGTTTTAATCAGCAGAGCATGTTTTTA	1276
DB	4335	TCCAAAGCATTAATATCCCAATCATGCGCTGATGTTTTAATCAGCAGAGCATGTTTTTA	4394
QY	1277	TGTTTGTACAAAGAGATTTATGGGTGGGATGGAGTATAGACCATGCAATGTCAC	1336
DB	4395	TGTTTGTACAAAGAGATTTATGGGTGGGATGGAGTATAGACCATGCAATGTCAC	4454
QY	1337	CTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGGACTCTGAACAGACACC	1396
DB	4455	CTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGGACTCTGAACAGACACC	4514
QY	1397	TAATAATGGGTTGATGCTGAAGTAGCAAACTCTTCTGGAACGCAAACTCTTTTAAGGAA	1456
DB	4515	TAATAATGGGTTGATGCTGAAGTAGCAAACTCTTCTGGAACGCAAACTCTTTTAAGGAA	4574
QY	1457	GTCCCTAATTTAGAAACCCCAACTTCAACATATCATTAATAGCAAACTTTGAAGG	1516
DB	4575	GTCCCTAATTTAGAAACCCCAACTTCAACATATCATTAATAGCAAACTTTGAAGG	4634
QY	1517	AAGTCTCTTGAATTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	1576
DB	4635	AAGTCTCTTGAATTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	4694

QY	1577	GAATGCCAATCAGGTCAGGTTTGCTACATTTTGTATGTGTGTGTGTGTCTCTCCCAAG	1636
DB	4695	GAATGCCAATCAGGTCAGGTTTGCTACATTTTGTATGTGTGTGTGTGTCTCTCCCAAG	4754
QY	1637	GTATATTAACATATATAGAGAGTTGTGACAAAACAGAAATGATAAAGTGGCAACCGTGC	1696
DB	4755	GTATATTAACATATATAGAGAGTTGTGACAAAACAGAAATGATAAAGTGGCAACCGTGC	4814
QY	1697	ACAGCTCATAGTCTTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTTGAACACAGT	1756
DB	4815	ACAGCTCATAGTCTTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTTGAACACAGT	4874
QY	1757	GTTCAAGGCCAGCTGGGCAACATACAAAGATCCTGTCTCAAAAAAATAAAAAA	1816
DB	4875	GTTCAAGGCCAGCTGGGCAACATACAAAGATCCTGTCTCAAAAAAATAAAAAA	4934
QY	1817	AAGAAAGAGAGAGGGCGGCGTGGTGTGCTACGCCCTGTAAATCCCAAGCACTTTGGGAGC	1876
DB	4935	AAGAAAGAGAGAGGGCGGCGTGGTGTGCTACGCCCTGTAAATCCCAAGCACTTTGGGAGC	4994
QY	1877	CGAGCCGGCGGATCACTGTGCTCAGGAGTTTGAACACCAAGCTGGCCCAACATGGCAAA	1936
DB	4995	CGAGCCGGCGGATCACTGTGCTCAGGAGTTTGAACACCAAGCTGGCCCAACATGGCAAA	5054
QY	1937	CCCGCTGTCTACTCAAAATGCAAAAAATTTAGCCAGGCTGTAGCAGGCACTCTTAATCCC	1996
DB	5055	CCCGCTGTCTACTCAAAATGCAAAAAATTTAGCCAGGCTGTAGCAGGCACTCTTAATCCC	5114
QY	1997	AGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTTGAACCCAGGAGTGGAGTTGCAGTA	2056
DB	5115	AGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTTGAACCCAGGAGTGGAGTTGCAGTA	5174
QY	2057	AGCTGAGATGCTGCGCTTGACCTCCAGCCTGGCGGACAGCAAGCAAGCTCTGTCTCAGAA	2116
DB	5175	AGCTGAGATGCTGCGCTTGACCTCCAGCCTGGCGGACAGCAAGCAAGCTCTGTCTCAGAA	5234
QY	2117	AAAAAATAAAG	2176
DB	5235	AAAAAATAAAG	5294
QY	2177	AAGCATTCGAGGAGAAATTTGTCTTTATCCAAACAAATGTAAGGAGGCAATTAAGGATCCC	2236
DB	5295	AAGCATTCGAGGAGAAATTTGTCTTTATCCAAACAAATGTAAGGAGGCAATTAAGGATCCC	5354
QY	2237	TATTTCTCTCTTTTGGTGTCTATTGTCCCTTAACAACAGTCTTTGACAGTGAGAGAAATA	2296
DB	5355	TATTTCTCTCTTTTGGTGTCTATTGTCCCTTAACAACAGTCTTTGACAGTGAGAGAAATA	5414
QY	2297	TTCAGAAATACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCATTAAGAGATGAGC	2356
DB	5415	TTCAGAAATACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCATTAAGAGATGAGC	5474
QY	2357	AGATCCACAGGAGAACTTGAATCCACAACTGTCTTTATTTAAATCTTATTTGATACATAGTT	2416
DB	5475	AGATCCACAGGAGAACTTGAATCCACAACTGTCTTTATTTAAATCTTATTTGATACATAGTT	5534
QY	2417	TGTAAGAGAGTTAAAAATTTCTTACTTCATGTTATTTATTTATTTATTTATTTTTCGCT	2476
DB	5535	TGTAAGAGAGTTAAAAATTTCTTACTTCATGTTATTTATTTATTTATTTATTTTTCGCT	5594
QY	2477	CTAATGATTTTATTAACATGATTTCTCTTCTGATATATTTGAAATGGAGTCTCAAGC	2536
DB	5595	CTAATGATTTTATTAACATGATTTCTCTTCTGATATATTTGAAATGGAGTCTCAAGC	5654
QY	2537	TTCAATAATTTTAACTTTAGAAATGATCTTAATAACAGAGTATGTAATTTGTAACATTC	2596
DB	5655	TTCAATAATTTTAACTTTAGAAATGATCTTAATAACAGAGTATGTAATTTGTAACATTC	5714
QY	2597	AGTAATGGTGTACGAGGCAATTTCTTGTATTTTGTAGTAACCTTTTATGACAGCAAAAT	2656
DB	5715	AGTAATGGTGTACGAGGCAATTTCTTGTATTTTGTAGTAACCTTTTATGACAGCAAAAT	5774

QY 2657 TGCCTTCTGGCTCACTTTCAATCAGTTAAATAATGATAAATAATTTTGAAGCTGTGAAG 2716
 Db 5775 TGCCTTCTGGCTCACTTTCAATCAGTTAAATAATGATAAATAATTTTGAAGCTGTGAAG 5834

QY 2717 ATAAATACCAATAAATAATATAATAAGTGATTTATATGAAGTTAAATAATAAAATCAG 2776
 Db 5835 ATAAATACCAATAAATAATATAATAAGTGATTTATATGAAGTTAAATAATAAAATCAG 5894

QY 2777 TATGATGAATAAACTTGAAA 2797
 Db 5895 TATGATGAATAAACTTGAGA 5915

RESULT 3
 AAC55339
 ID AAC55339 standard; DNA; 11204 BP.
 XX
 AC AAC55339;
 XX
 DT 05-FEB-2001 (first entry)
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; antiallergic;
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
 KW
 OS Homo sapiens.
 XX
 PN WO200058480-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 28-MAR-2000; 2000WO-JP01918.
 XX
 PR 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX
 PA (NISB) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX
 PI Honjo T, Muramatsu M;
 XX
 DR WPI; 2000-611715/58.
 XX
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 PS Claim 17; Page 163-170; 174pp; Japanese.
 XX
 CC The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, Digeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences

CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents a genomic DNA sequence of human AID.
 XX
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;
 Query Match 77.2%; Score 2174.6; DB 21; Length 11204;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 617 CTTTGGCCCTCTATGAGTTGATGACTTTACAGAGCGCATTTCTGACTTTGGGACTTTGA 676
 Db 8950 CTACAGCCCTCTATGAGTTGATGACTTTACAGACGCGCATTTCTGACTTTGGGACTTTGA 9009

QY 677 TAGCAACTTCCAGGAATGTCACACAGCATGAATAATCTCTGCTGAAGACAGTGAATAAA 736
 Db 9010 TAGCAACTTCCAGGAATGTCACACAGCATGAATAATCTCTGCTGAAGACAGTGAATAAA 9069

QY 737 AACAGTCTCTCAAGTCTTCTCTGTTTATTTCTCAACTCTCTCAGTTTCTTAGAGTTTACA 796
 Db 9070 AACAGTCTCTCAAGTCTTCTCTGTTTATTTCTCAACTCTCTCAGTTTCTTAGAGTTTACA 9129

QY 797 GAAAAATATTTATATAGACTCTTTTAAAGATCTATGCTTTGAAATAGAGAAGAAC 856
 Db 9130 GAAAAATATTTATATAGACTCTTTTAAAGATCTATGCTTTGAAATAGAGAAGAAC 9189

QY 857 ACAGTCTGGCCAGGAGCTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 916
 Db 9190 ACAGTCTGGCCAGGAGCTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 9249

QY 917 CTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCTATGA 976
 Db 9250 CTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCTATGA 9309

QY 977 CTTTATAGTAGGATGAGACCAAGTAGATCCTTAAAGACATGCTGAGAGGATCAATG 1036
 Db 9310 CTTTATAGTAGGATGAGACCAAGTAGATCCTTAAAGACATGCTGAGAGGATCAATG 9369

QY 1037 TTTTATATCAACATCCTTTTATTTGATTTGAGTTTAACTGAGTGTGTAGTGATA 1096
 Db 9370 TTTTATATCAACATCCTTTTATTTGATTTGAGTTTAACTGAGTGTGTAGTGATA 9429

QY 1097 GATTTTCTATTTCTTTTCCCTTGAGCTTTTCAAGTAAACACAACTCTTCCATCAGG 1156
 Db 9430 GATTTTCTATTTCTTTTCCCTTGAGCTTTTCAAGTAAACACAACTCTTCCATCAGG 9489

QY 1157 CCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTTGTGACCCCAACCATCTC 1216
 Db 9490 CCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTTGTGACCCCAACCATCTC 9549

QY 1217 TCCAAAGCATTAATATCCCAATCATGCGTGTATGTTTAACTCAGACAGAGCATGTTTTTA 1276
 Db 9550 TCCAAAGCATTAATATCCCAATCATGCGTGTATGTTTAACTCAGACAGAGCATGTTTTTA 9609

QY 1277 TGTTTGTACAAAAGAGATTGTTATGGGTGGGATGGAGGTATAGACCATGATGTCAC 1336
 Db 9610 TGTTTGTACAAAAGAGATTGTTATGGGTGGGATGGAGGTATAGACCATGATGTCAC 9669

QY 1337 CTTCAAGCTACTTTTAAATAAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACC 1396
 Db 9670 CTTCAAGCTACTTTTAAATAAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACC 9729

QY 1397 TAATATGGGTGATGTCCTGAAGTAGCAAAATCTCTGGAACGCAAACTCTTTTAAAGGA 1456
 Db 9730 TAATATGGGTGATGTCCTGAAGTAGCAAAATCTCTGGAACGCAAACTCTTTTAAAGGA 9789

QY 1457 GTCCCTAATTTAGAACACCCCAAACTTCAATATCATATATAGCAAACTTTGAAGG 1516
 Db 9790 GTCCCTAATTTAGAACACCCCAAACTTCAATATCATATATAGCAAACTTTGAAGG 9849

QY 1517 AAGTTGCTTGAATGTTGGGAGGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA 1576
 Db 9850 AAGTTGCTTGAATGTTGGGAGGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA 9909

QY 1577 GAAATGCAATCAGGTCAAGTGTGCTACATTTTGTATGCTGTGATGCTTCTCCCAAAG 1636
DB 9910 GAAATGCCATCAGGTCAAGTGTGCTACATTTTGTATGCTGTGATGCTTCTCCCAAAG 9969
QY 1637 GTATATTAATATATAAGAGAGTTGTGCACAAACAGAAATGATAAAGCTGCGAACCGTGGC 1696
DB 9970 GTATATTAATATATAAGAGAGTTGTGCACAAACAGAAATGATAAAGCTGCGAACCGTGGC 10029
QY 1697 ACAGGCTCATAGTTCTAGCTGCTTGGAGAGTTGAGGAGGAGGATGCTTGAACACAGGT 1756
DB 10030 ACAGGCTCATAGTTCTAGCTGCTTGGAGAGTTGAGGAGGAGGATGCTTGAACACAGGT 10089
QY 1757 GTTCAAGCCGAGCTGGGCAACATACAAAGATCTGCTCTCAAAAAAATAAAAAA 1816
DB 10090 GTTCAAGCCGAGCTGGGCAACATACAAAGATCTGCTCTCAAAAAAATAAAAAA 10149
QY 1817 AAGAAAGAGAGAGCGCGGGCGTGGTGGCTACAGCCCTGTATCCACAGCATTGGGAGGC 1876
DB 10150 AAGAAAGAGAGAGCGCGGGCGTGGTGGCTACAGCCCTGTATCCACAGCATTGGGAGGC 10209
QY 1877 CGAGCCGGCGGATCAGCTGTGCTCAGGAGTTTGAGACCAAGCTTGGCCACATGGCAAAA 1936
DB 10210 CGAGCCGGCGGATCAGCTGTGCTCAGGAGTTTGAGACCAAGCTTGGCCACATGGCAAAA 10269
QY 1937 CCGCGTCTGTACTCAAAATGCAAAATAGCCAGCGTGTAGCAGGCACCTGTGTAATCCC 1996
DB 10270 CCGCGTCTGTACTCAAAATGCAAAATAGCCAGCGTGTAGCAGGCACCTGTGTAATCCC 10329
QY 1997 AGCTACTTGGAGGCTGAGCAGAGAGTAATCGTTGAACCCAGGAGTGGAGTTGTCAGTA 2056
DB 10330 AGCTACTTGGAGGCTGAGCAGAGAGTAATCGTTGAACCCAGGAGTGGAGTTGTCAGTA 10389
QY 2057 AGCTGAGATCGTCCGCTGTCACCTCCAGCCTGGCGCAAGCAAGACTCTGCTCAGAA 2116
DB 10390 AGCTGAGATCGTCCGCTGTCACCTCCAGCCTGGCGCAAGCAAGACTCTGCTCAGAA 10449
QY 2117 AAAAAAAG 2176
DB 10450 AAAAAAAG 10509
QY 2177 AAGCATTCGAAGAAATTTGCTTTATCCAAACAAATGTAAAGAGCCCAATGAAGGATCCC 2236
DB 10510 AAGCATTCGAAGAAATTTGCTTTATCCAAACAAATGTAAAGAGCCCAATGAAGGATCCC 10569
QY 2237 TATTTGCTCTTTTGGTGTCTATTGTGCTTACCAACTGCTTTGACAGTGAGAAATA 2296
DB 10570 TATTTGCTCTTTTGGTGTCTATTGTGCTTACCAACTGCTTTGACAGTGAGAAATA 10629
QY 2297 TTCAGATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCAATGAAGATGAGC 2356
DB 10630 TTCAGATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCAATGAAGATGAGC 10689
QY 2357 AGATCCACAGAAACTTGAATGCACACACTGCTTATTTTAATCTTATTTGATACATAGTT 2416
DB 10690 AGATCCACAGAAACTTGAATGCACACACTGCTTATTTTAATCTTATTTGATACATAGTT 10749
QY 2417 TGTAAAGAGCTTAAATTTCTTACTCATGCTATTTCATTTATTTATTTATTTTTCGCT 2476
DB 10750 TGTAAAGAGCTTAAATTTCTTACTCATGCTATTTCATTTATTTATTTATTTTTCGCT 10809
QY 2477 CTAATGATTTTATTAACATGATTTCCCTTTCTGATATATTGAAATGGAGTCTCAAGC 2536
DB 10810 CTAATGATTTTATTAACATGATTTCCCTTTCTGATATATTGAAATGGAGTCTCAAGC 10869
QY 2537 TCCATTAATTTTAACTTTAGAATGATTTCTATATACACAGTATGTAATTTGATACATTC 2596
DB 10870 TCCATTAATTTTAACTTTAGAATGATTTCTATTAACACAGTATGTAATTTGATACATTC 10929
QY 2597 AGTAATGGTCTACGAGCATTCTCTTGTATTTTGTAACTTTTATGACAGCAAAAT 2656
DB 10930 AGTAATGGTCTACGAGCATTCTCTTGTATTTTGTAACTTTTATGACAGCAAAAT 10989

QY 2657 TGCTTCTGGCTCACTTTCAATCACTTAAATAAATGATAAATAATTTTGGAGCTGTGAAG 2716
DB 10990 TGCTTCTGGCTCACTTTCAATCACTTAAATAAATGATAAATAATTTTGGAGCTGTGAAG 11049
QY 2717 ATAAATACCAAAATAAATAAATAAAGTGATTATATGAAGTTAAATAAATAAATCAG 2776
DB 11050 ATAAATACCAAAATAAATAAATAAAGTGATTATATGAAGTTAAATAAATAAATCAG 11109
QY 2777 TATGATGGAATAAATTTGAAA 2797
DB 11110 TATGATGGAATAAATTTGAGA 11130
RESULT 4
AAC55319
ID AAC55319 standard; DNA: 2172 BP.
XX AC AAC55319;
XX AC
XX AC
DT 05-FEB-2001 (first entry)
DE Human activation-induced cytidine deaminase exon 5 SEQ ID NO:15.
XX
XX
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
immune related disease; allergy; allergic disease; antiallergic;
antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
gene therapy; B cell associated immune system disorder; food allergy;
immunodeficiency disease; immunoglobulin A deficiency disease; aschma;
IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
Kw allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
ataxia telangiectasia; common variable immunodeficiency disorder;
Kw major histocompatibility class II deficiency disease;
Kw auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
XX Homo sapiens.
OS
XX WO200058480-A1.
XX
XX
PD 05-OCT-2000.
XX
PF 28-MAR-2000; 2000WO-JP01918.
XX
XX 29-MAR-1999; 99JP-0087192.
PR 24-JUN-1999; 99JP-0178999.
PR 27-DEC-1999; 99JP-0371382.
XX
XX (NISR) JAPAN TOBACCO INC.
PA (HONJ) HONJO T.
XX
XX Honjo T, Muramatsu M;
XX WPI: 2000-611715/58.
XX
XX
PT Nucleic acid encoding activation induced cytidine deaminase, useful as
PT a target for drug development for immune-related diseases including
PT allergies -
XX
XX
PS Claim 18; Page 152-153; 174pp; Japanese.
XX
XX
CC The present invention describes an activation-induced cytidine deaminase
(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
CC dermatological activities, and can be used in gene therapy. AID
CC polynucleotides are useful in methods for identifying drugs for the
CC treatment of B cell associated immune system disorders, immunodeficiency
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic
CC colitis, aschma, food allergy, drug allergy, allergic rhinitis, Rosen
CC disease, Digeorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class) class
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences

CC encoding AID may be used for gene therapy and the antibodies to the AID
CC protein may be used for diagnosis and treatment of these disorders. The
CC present sequence represents the exon 5 genomic DNA sequence of human AID.
XX
SQ Sequence 2172 BP; 702 A; 379 C; 465 G; 626 T; 0 other;

QY	2663	TGCTCAGTTCATCAATCAAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAGATAAAA	2722
Db	2041	TGGCTCAGTTCATCAATCAAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAGATAAAA	2100
QY	2723	TACCAATAAATAAATAAATAAAGTATTTATGAACTTAAATAAATAAATAAATCAGTATGAT	2782
Db	2101	TACCAATAAATAAATAAATAAAGTATTTATGAACTTAAATAAATAAATAAATCAGTATGAT	2160
QY	2783	GGAATAAACTTG 2794	
Db	2161	GGAATAAACTTG 2172	
RESULT 5			
AAK81088			
ID	AAK81088 standard; DNA; 1665 BP.		
XX	AAK81088;		
XX	07-NOV-2001 (first entry)		
XX	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35900.		
DE	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;		
KW	cytostatic; gene therapy; vaccine; metastasis; ds.		
XX	Homo sapiens.		
XX	W0200157182-A2.		
PN	09-AUG-2001.		
PD	17-JAN-2001; 2001WO-US01354.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180828.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	18-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226868.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 2000US-0235836.		
PR	29-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236367.		
PR	29-SEP-2000; 2000US-0236368.		
PR	29-SEP-2000; 2000US-0236369.		
PR	29-SEP-2000; 2000US-0236370.		
PR	02-OCT-2000; 2000US-0236802.		
PR	02-OCT-2000; 2000US-0237037.		
PR	02-OCT-2000; 2000US-0237038.		
PR	02-OCT-2000; 2000US-0237039.		
PR	13-OCT-2000; 2000US-0237040.		
PR	13-OCT-2000; 2000US-0239935.		
PR	13-OCT-2000; 2000US-0239937.		
PR	20-OCT-2000; 2000US-0240960.		
PR	20-OCT-2000; 2000US-0241221.		
PR	20-OCT-2000; 2000US-0241785.		
PR	20-OCT-2000; 2000US-0241786.		
PR	20-OCT-2000; 2000US-0241787.		
PR	20-OCT-2000; 2000US-0241808.		
PR	20-OCT-2000; 2000US-0241809.		
PR	20-OCT-2000; 2000US-0241826.		
PR	01-NOV-2000; 2000US-0244617.		
PR	08-NOV-2000; 2000US-0246474.		
PR	08-NOV-2000; 2000US-0246475.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246478.		
PR	08-NOV-2000; 2000US-0246523.		
PR	08-NOV-2000; 2000US-0246524.		
PR	08-NOV-2000; 2000US-0246525.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246527.		
PR	08-NOV-2000; 2000US-0246528.		
PR	08-NOV-2000; 2000US-0246532.		
PR	08-NOV-2000; 2000US-0246610.		
PR	08-NOV-2000; 2000US-0246611.		
PR	08-NOV-2000; 2000US-0246613.		
PR	17-NOV-2000; 2000US-0249207.		
PR	17-NOV-2000; 2000US-0249208.		
PR	17-NOV-2000; 2000US-0249209.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249211.		
PR	17-NOV-2000; 2000US-0249212.		
PR	17-NOV-2000; 2000US-0249213.		
PR	17-NOV-2000; 2000US-0249214.		

PR	17-NOV-2000;	2000US-0249215.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
----	--------------	-----------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Db 1432 CGAGCCGGGGGATCACTGTGGTCAGAGTTTGAGACCAAGCCTGGCCCAACATGCAAAA 1491
 Qy 1937 CCCCCTGTCTACTCAAAATGCAAAATTTAGCCAGCGCTGGTAGCAGGCACCTGTAATCCC 1996
 Db 1492 CCCCCTGTCTACTCAAAATGCAAAATTTAGCCAGCGCTGGTAGCAGGCACCTGTAATCCC 1551
 Qy 1997 AGCTACTGGAGGCTGAGGAGGAGAGATCGCTTTGAACCCAGGAGGTGGAGTTGCAGTA 2056
 Db 1552 AGCTACTGGAGGCTGAGGAGGAGAGATCGCTTTGAACCCAGGAGGTGGAGTTGCAGTA 1611
 Qy 2057 AGCTGAGATGCGCGCTGACCTCCAGCCTGGCGGCAAGAGCAAGACTCTGTC 2110
 Db 1612 AGCTGAGATGCGCGCTGACCTCCAGCCTGGCGGCAAGAGCAAGACTCTGTC 1665

RESULT 6

AAC55307

ID AAC55307 standard; cDNA; 2440 BP.

XX AAC55307;

DT 05-FEB-2001 (first entry)

DE Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.

KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; antiallergic;
 KW antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease;
 KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.

XX Mus musculus.

XX Key Location/Qualifiers

FH 93..689

FT CDS /*tag= a

FT /product= "activation-induced cytidine deaminase"

XX W0200058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

PR 27-DEC-1999; 99JP-0371382.

XX (NIBS) JAPAN TOBACCO INC.

PA (HONJ/) HONJO T.

XX Honjo T, Muramatsu M;

XX WPI; 2000-611715/58.

XX P-PSDB; AAB24197.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as

XX a target for drug development for immune-related diseases including

XX allergies -

XX Claim 3; Page 126-130; 174pp; Japanese.

XX The present sequence encodes mouse activation-induced cytidine deaminase

XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and

XX has cytidine activity similar to APOBEC-1. AID has antiallergic,

XX antianemic, antisthmatic, ophthalmological, anti-HIV and

XX dermatological activities, and can be used in gene therapy. AID

CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (Iga) deficiency
 CC disease, Iga nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class)
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC Ige disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.

XX SQ Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 1 other;

Query Match 21.4%; Score 603.4; DB 21; Length 2440;

Best Local Similarity 69.4%; Pred. No. 2.4e-91;

Matches 1006; Conservative 0; Mismatches 351; Indels 92; Gaps 10;

Qy 4 GAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTCGAGGGAGCAAG-AAAG 62

Db 16 GAAGCAGCCTTGCTTGAAGCAAGCTTCCTTTGGCCTAAGACTTTTGAGGAGCTCAAGAAAG 75

Qy 63 ACACCTCTGGACACCACTATGACAGACCTCTTGATGAACCGGAGGAAGTTCTTTACCAAT 122

Db 76 TCAGCCTGGAGACCGATATGACAGCCTTCTGATGAAGCAAAAGAAGTTTCTTTACCAAT 135

Qy 123 TCAAAATGTCCGCTGGCTAAGGTGCGGTGAGACCTACCTGCTGCTACCTAGTGAAGA 182

Db 136 TCAAAATGTCCGCTGGCCTGAGGACGCGATGAGACCTACCTGCTGCTGCTGCTGGAAGA 195

Qy 183 GCGGTGACAGTGCTACATCCTTTTCTGCTGACCTTTGGTATTTTTCGCAATAAGAACCGCT 242

Db 196 GGAGAGATAGTGCCACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 255

Qy 243 GCCACGTGGAAATGCTCTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302

Db 256 GCCACGTGGAAATGCTCTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 315

Qy 303 GCTACCGCGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362

Db 316 GTTACCGCGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375

Qy 363 CCGACTTTCTGCGAGGAAACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422

Db 376 CTGAGTTTCTGAGATGGAACCTTAACCTGACCTGAGGATTTTTCACCGCGCTGCTGCTGCT 435

Qy 423 TCTGTGAGGACCGCAAGGCTGACCGCGAGGCGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 482

Db 436 TCTGTGAAGACCGCAAGGCTGAGCGCTGAGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 495

Qy 483 AAATAGCATCATGACCTTCAAGATTTATTTTACTGCTGGAATACTTTTGTAGAAACCC 542

Db 496 AGATCGGATCATGACCTTCAAGACTATTTTACTGCTGGAATACATTTGTAGAAATC 555

Qy 543 ATGAAAGAACTTTCAAGCGCTGGGAAGGCTGCAATGAAATTCAGTTGCTGCTCTCCAGAC 602

Db 556 GTGAAAGAACTTTCAAGCGCTGGGAAGGCTGCAATGAAATTCAGTTGCTGCTCTCCAGAC 615

Qy 603 AGCTTCGGCGCATCCTTTTGGCCCTGATGAGTTGATGACTTACGAGAGCATTTTCGTA 662

Db 616 AACTTCGGCGCATCCTTTTGGCCCTGATGAGTTGATGACTTACGAGAGCATTTTCGTA 675

Qy 663 CTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACAGATGAAATATCTCTGCTGAA 722

Db 676 TGTGGGATTTGAAAGCAACCTCCTGGAATGTCACACGATGATGAAATTTCT---CTGAA 732

Qy 723 GACATGGATGAAAGACAGTCTTCAAGTCTTCTGCTTTTATTTTATTTCAACTCTCACTT 782

Db 733 GAGATGGATGAGAAAGCAACCTTCA--ACTACATGTTTCTTCTTCTTAAGTACTCACTT 790

Qy 783 TCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAGAGATCTATGCTGTA 842

Db 791 TTATAAGTGTAGGGGAAA-----TTATATGACTTTTAAAAAATACTTTGAGCTGCAC 843

PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251388.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.

PA
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX

PS Disclosure; SEQ ID NO 35901; 3071pp + Sequence Listing; English.
 XX
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC

CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 574 BP; 201 A; 80 C; 85 G; 208 T; 0 other;

Query Match 20.4%; Score 574; DB 22; Length 574;
 Best Local Similarity 100.0%; Pred. No. 1.6e-86;
 Matches 574; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2211 AATGTAAGGAGCCCAATAAGGGATCCCTATTGCTCTCTTTGGTGTCTATTGTCCTTAAC 2270
 Db 1 AATGTAAGGAGCCCAATAAGGGATCCCTATTGCTCTCTTTGGTGTCTATTGTCCTTAAC 60
 QY 2271 AACTGTCTTTGACAGTGAGAAAAATATTCAGATAACCATATCCCTGTGCCGTTATTACC 2330
 Db 61 AACTGTCTTTGACAGTGAGAAAAATATTCAGATAACCATATCCCTGTGCCGTTATTACC 120
 QY 2331 TAGCAACCCCTTGCATGAAGATGAGCAGATCCACAGAAACCTTGAATGCACAACTGTCT 2390
 Db 121 TAGCAACCCCTTGCATGAAGATGAGCAGATCCACAGAAACCTTGAATGCACAACTGTCT 180
 QY 2391 TATTTTAATCTTATTGTACATAAGTTTGTAAAGAGCTTAAAAATTTCTTACTTCATCTATT 2450
 Db 181 TATTTTAATCTTATTGTACATAAGTTTGTAAAGAGCTTAAAAATTTCTTACTTCATCTATT 240
 QY 2451 CATTTATATTTTATATTTTTCGCTCTAATGATTTTATTAACATGATTTCTCTTTTCT 2510
 Db 241 CATTTATATTTTATATTTTTCGCTCTAATGATTTTATTAACATGATTTCTCTTTTCT 300
 QY 2511 GATATATGAATGGAGTCTCAAGCTTCATAAAATTTATAAATTTAGAAATGATTTCTAAT 2570
 Db 301 GATATATGAATGGAGTCTCAAGCTTCATAAAATTTATAAATTTAGAAATGATTTCTAAT 360
 QY 2571 AACACGTATGTAATGTAACATTTGCAGTAATGGTCTACGAGGCCATTTCTCTCATTT 2630
 Db 361 AACACGTATGTAATGTAACATTTGCAGTAATGGTCTACGAGGCCATTTCTCTCATTT 420
 QY 2631 TTAGTAAACTTTTATGACAGCAAAATTTGCTTCTGCTCACTTTTCAATCAGTTAAATAAAT 2690
 Db 421 TTAGTAAACTTTTATGACAGCAAAATTTGCTTCTGCTCACTTTTCAATCAGTTAAATAAAT 480
 QY 2691 GATAATAAATTTTGGAGCTGTGAAGATAAATACCAATAAATAAATAAATAAATGATTT 2750
 Db 481 GATAATAAATTTTGGAGCTGTGAAGATAAATACCAATAAATAAATAAATAAATGATTT 540
 QY 2751 TATATGAAGTTAAAAATAAAAAATCAGTATGATGG 2784
 Db 541 TATATGAAGTTAAAAATAAAAAATCAGTATGATGG 574

RESULT 8
 AAK61819
 ID AAK61819 standard; cDNA; 429 BP.
 XX
 AC AAK61819;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/hematopoietic antigen encoding cDNA SEQ ID NO:6879.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
DR P-PSDB; AAM89038.
XX

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Claim 1; SEQ ID NO 6879; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.

XX SQ Sequence 429 BP; 144 A; 79 C; 97 G; 109 T; 0 other;

Query Match 15.2%; Score 429; DB 22; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-62;
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1386 ACAAGACACCTAATATGGTTGATGCTCTGAAGTAGCAATCTCTGGAACGCAACT 1445
DB 1 ACAAGACACCTAATATGGTTGATGCTCTGAAGTAGCAATCTCTGGAACGCAACT 60
QY 1446 CTTTAAAGGAAGTCCCTAATTTAGAACACCCCAAACTTCACATATCATATTAAGCAAA 1505
DB 61 CTTTAAAGGAAGTCCCTAATTTAGAACACCCCAAACTTCACATATCATATTAAGCAAA 120
QY 1506 CAATTGGAAGGAAGTTCCTGTAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTC 1565
DB 121 CAATTGGAAGGAAGTTCCTGTAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTC 180
QY 1566 TCCTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCATATTTTGTATGTGTGATGC 1625
DB 181 TCCTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCATATTTTGTATGTGTGATGC 240
QY 1626 TTCTCCCAAGGTATATTAAGTATATAAGAGAGTTGTGACAAACAGAAATGATAAAGCTG 1685
DB 241 TTCTCCCAAGGTATATTAAGTATATAAGAGAGTTGTGACAAACAGAAATGATAAAGCTG 300
QY 1686 CGAACCGTGCACACGCTCATAGTCTTACGTTGGGAGGTTGAGGAGGAGATGGCT 1745
DB 301 CGAACCGTGCACACGCTCATAGTCTTACGTTGGGAGGTTGAGGAGGAGATGGCT 360
QY 1746 TGAACACAGGTGTTCAAGGCCACCTGGGCAACATACAGATCTCTCTCAAAAAAAA 1805
DB 361 TGAACACAGGTGTTCAAGGCCACCTGGGCAACATACAGATCTCTCTCAAAAAAAA 420
QY 1806 AAAAAAAA 1814
DB 421 AAAAAAAA 429

RESULT 9
ABA08208

ID ABA08208 standard; DNA; 13862 BP.
XX
AC ABA08208;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 1003.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilucer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ds.
XX
OS Homo sapiens.
XX
PN WO200155325-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01345.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.

PR	12-SEP-2000;	2000US-0231968
PR	14-SEP-2000;	2000US-0232397
PR	14-SEP-2000;	2000US-0232398
PR	14-SEP-2000;	2000US-0232399
PR	14-SEP-2000;	2000US-0232400
PR	14-SEP-2000;	2000US-0232401
PR	14-SEP-2000;	2000US-0232403
PR	14-SEP-2000;	2000US-0233064
PR	14-SEP-2000;	2000US-0233065
PR	21-SEP-2000;	2000US-0232423
PR	21-SEP-2000;	2000US-0232427
PR	25-SEP-2000;	2000US-0234997
PR	25-SEP-2000;	2000US-0234998
PR	26-SEP-2000;	2000US-0235484
PR	27-SEP-2000;	2000US-0235834
PR	29-SEP-2000;	2000US-0235836
PR	29-SEP-2000;	2000US-0236327
PR	29-SEP-2000;	2000US-0236367
PR	29-SEP-2000;	2000US-0236368
PR	29-SEP-2000;	2000US-0236369
PR	29-SEP-2000;	2000US-0236370
PR	02-OCT-2000;	2000US-0236802
PR	02-OCT-2000;	2000US-0237037
PR	02-OCT-2000;	2000US-0237038
PR	02-OCT-2000;	2000US-0237039
PR	13-OCT-2000;	2000US-0237040
PR	13-OCT-2000;	2000US-0239935
PR	13-OCT-2000;	2000US-0239937
PR	20-OCT-2000;	2000US-0241187
PR	20-OCT-2000;	2000US-0241809
PR	20-OCT-2000;	2000US-0241826
PR	01-NOV-2000;	2000US-0244617
PR	08-NOV-2000;	2000US-0246474
PR	08-NOV-2000;	2000US-0246475
PR	08-NOV-2000;	2000US-0246476
PR	08-NOV-2000;	2000US-0246477
PR	08-NOV-2000;	2000US-0246478
PR	08-NOV-2000;	2000US-0246523
PR	08-NOV-2000;	2000US-0246524
PR	08-NOV-2000;	2000US-0246525
PR	08-NOV-2000;	2000US-0246526
PR	08-NOV-2000;	2000US-0246527
PR	08-NOV-2000;	2000US-0246532
PR	08-NOV-2000;	2000US-0246609
PR	08-NOV-2000;	2000US-0246610
PR	08-NOV-2000;	2000US-0246611
PR	08-NOV-2000;	2000US-0246613
PR	17-NOV-2000;	2000US-0249207
PR	17-NOV-2000;	2000US-0249208
PR	17-NOV-2000;	2000US-0249209
PR	17-NOV-2000;	2000US-0249210
PR	17-NOV-2000;	2000US-0249211
PR	17-NOV-2000;	2000US-0249212
PR	17-NOV-2000;	2000US-0249213
PR	17-NOV-2000;	2000US-0249214
PR	17-NOV-2000;	2000US-0249215
PR	17-NOV-2000;	2000US-0249216
PR	17-NOV-2000;	2000US-0249217
PR	17-NOV-2000;	2000US-0249218
PR	17-NOV-2000;	2000US-0249244
PR	17-NOV-2000;	2000US-0249245
PR	17-NOV-2000;	2000US-0249264
PR	17-NOV-2000;	2000US-0249265
PR	17-NOV-2000;	2000US-0249297
PR	17-NOV-2000;	2000US-0249299
PR	01-DEC-2000;	2000US-0249300
PR	01-DEC-2000;	2000US-0250160

PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-)	HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-498786/53.		
XX	New isolated ovarian and/or breast cancer related nucleic acids and		
PT	polypeptides, useful for diagnosing, treating and/or preventing human		
PT	diseases and disorders, particularly ovarian and/or breast cancer -		
XX	Disclosure; SEQ ID NO 1003; 577pp + Sequence Listing; English.		
XX	The invention relates to novel genes (ABA07454-ABA08224) and proteins		
CC	(ABB10743-ABB10980) useful for preventing, treating or ameliorating		
CC	medical conditions e.g. by protein or gene therapy. The genes are		
CC	isolated from a range of human tissues disclosed in the specification.		
CC	The nucleic acids, proteins, antibodies and antagonists are useful		
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast		
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone		
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;		
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune		
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's		
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative		
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;		
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and		
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal		
CC	and parasitic infections.		
CC	Note: The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 other;		
Query Match	10.58; Score 295.2; DB 22; Length 13862;		
Best Local Similarity	74.6%; Pred. No. 3.7e-40;		
Matches 397; Conservative	0; Mismatches 133; Indels 2; Gaps 2;		
Qy	1667	AAACAGAAATGATAAAGCTGCGAACCGGTGCACACAGCTCATAGTTCTAGCTGCTTGGCGAGG	1736
Db	1508	AAACTTAATAGCTGGCTGGGCTCTGTTGGCTCATGCTGTGATCCCAAGCACTTGGGAGG	1567
Qy	1727	TTGAGGAGGAGGATGGCTTGAACACACAGTGTTTCAAGGCCAGCGCTGGGCAACATPAACAAG	1786
Db	1568	CTGAGGCAGGAGGATCACTTGAGTTCAGGAGTTTGAGACCAGCGCTGCCAACATGCGCAA	1627
Qy	1787	ATCCTGTCTCTCAAAAAAATAAAAAAAGAAAGA-GAGAGGGCCGGCGTGTGGTGGC	1845
Db	1628	ACCCCATCTCTACTTAAAAATACAAAAGTTTAGCCAGGCATGGCCGGCGATGGTGGGTGGC	1687
Qy	1846	TCAGCCCTGTAATCCCAAGCACTTTGGGAGCCGCGAGCGCGGATACCTGTGGTCAAGGA	1905
Db	1688	TCATGGCTGTAAATCCCAAGCACTTTGGGAGCCGAGGTGGCGGATACCTTGAGGTCAAGGA	1747
Qy	1906	GTTTGTGAGCACCAGCTGGCCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTA	1965
Db	1748	GTTCAAGACCAGCTGGCAACATGGCAAAACCCCTGTCTCTATTAAAAATACAAAATTA	1807
Qy	1966	GCCAGGCGTGGTAGCAGGCACCTGTAAATCCCAAGTACTTTGGGAGGCTTGAGCAGGAGAT	2025
Db	1808	GCCAGGCTTTATGTATGGCGGCTTAATCCAGTACTTCGAAGGCTTGAGCAGGAGAT	1867

```
QY 2026 CGCTGAACCCAGGAGGTGAGGTTGCAGTAAGCTGAGATCGTGCGCTTGCACCTCCAGCC 2085
Db 1868 CGTTGAAGCGAGGAGCGGAGTTTGCACGAGCCGAGATTGTGCCACTGCACCTCCAGCC 1927
QY 2086 TGGCGCACAGACCAAGACTCTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAA 2145
Db 1928 TGGCGCATAA-ACGAGACTCTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAA 1986
QY 2146 AGAGAAATATTTGGGAGAGAGAGGATGGGGAAGCATTTGCAAGGAATTTGTG 2197
Db 1987 AGAAAAATTAGCCAGCGTGGTGGCATGCACCTGTAGTCTAGCTACTTGGG 2038

RESULT 10
AAL02789
ID AAL02789 standard; DNA; 13862 BP.
XX AC AAL02789;
XX DT 21-NOV-2001 (first entry)
XX DE Human reproductive system related antigen DNA SEQ ID NO: 5477.
XX KW Human reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy; ds.
XX OS Homo sapiens.
XX PN W0200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 14-JUL-2000; 2000US-0217496.
XX PR 26-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 14-AUG-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 23-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226868.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231142.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231144.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
```


PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	08-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2000US-0254097.
XX		2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
XX	WPI; 2001-465570/50.	
DR		
XX		
PT	Isolated nucleic acid molecule e	
PT	is used in preventing, treating	
XX		
PS	Disclosure; SEQ ID NO 5477; 1297	
XX		
XX		
CC	The present invention provides t	
CC	number of human reproductive sys	
CC	in the prevention and treatment	
CC	including cancer. The present se	
CC	protein of the invention.	
XX		
SQ	Sequence 13862 BP; 3188 A; 3560	
	Query Match	10.5%; Score
	Best Local Similarity	74.6%; Pre
	Matches 397; Conservative	0;
QY	1667 AACACGAATGATAAGCTCGAA	
Db	1508 AACTTAATTAAGCTGCTGGGCT	
QY	1727 TTGAGGAGGAGGATGGCTTTGAA	
Db	1568 CTGAGCAGSAGGACTCACTTGAG	
QY	1787 ATCTGCTCTCTCAAAAAA	
Db	1628 ACCCCATCTCTACTAAAAATACA	
QY	1846 TCACGGCTCTAATCCAGCACATT	
Db	1688 TCATGGCTCTAATCCAGCACATT	
QY	1906 GTTTTGAGACCGAGCTGGCCAAAC	
Db	1748 GTTCAAGACCGAGCTGGCAACA	
QY	1966 GCCAGCGGTGTAGCAGGACACTT	
Db	1808 GCCAGCGTTTATGGTAGGCGCTT	

Qy	2026	CGTTGAACCCAGAGAGGTGGAGGTTGCAGTTAAGCTGAGATCGTGCGGTTGCACCTCCAGCC	2081
Db	1868	CGTTTGAAGCCAGGAGCGAGTTTGCACCGAGCCGAGATTTGTGCCACTGCACCTCCAGCC	1927
Qy	2086	TGGGCGACAGACGACACTCTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAA	2145
Db	1928	TGGGCGATAT-ACGAGAGACTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAA	1986
Qy	2146	AGAGAACATATTTGGGAGAGAGAGGATGGGGAAGCATTTGCAAGGAAATTTGTG	2197
Db	1987	AGAAAAATTAGCCACCGTGGTGCATGCACCTGTAGTCCCTAGCTACTTTGGG	2038
RESULT 11			
AAL07516			
XX	ID	AAL07516 standard; DNA; 13862 BP.	
AC	AAL07516;		
XX	XX	21-NOV-2001 (first entry)	
XX	XX	Human reproductive system related antigen DNA SEQ ID NO: 10204.	
XX	XX	Human; reproductive system related antigen; reproductive system disorder;	
KW	KW	cancer; gene therapy; ds.	
XX	XX	Homo sapiens.	
OS	OS	WO200155320-A2.	
PN	PN	02-AUG-2001.	
PD	PD	17-JAN-2001; 2001WO-US01339.	
XX	XX	31-JAN-2000; 2000US-0179065.	
XX	XX	04-FEB-2000; 2000US-0180628.	
PR	PR	24-FEB-2000; 2000US-0184564.	
PR	PR	02-MAR-2000; 2000US-0186350.	
PR	PR	16-MAR-2000; 2000US-0189874.	
PR	PR	17-MAR-2000; 2000US-0190076.	
PR	PR	18-APR-2000; 2000US-0198123.	
PR	PR	19-MAY-2000; 2000US-0205515.	
PR	PR	07-JUN-2000; 2000US-0209467.	
PR	PR	28-JUN-2000; 2000US-0214886.	
PR	PR	30-JUN-2000; 2000US-0215135.	
PR	PR	07-JUL-2000; 2000US-0216647.	
PR	PR	07-JUL-2000; 2000US-0216880.	
PR	PR	11-JUL-2000; 2000US-0217487.	
PR	PR	11-JUL-2000; 2000US-0217496.	
PR	PR	14-JUL-2000; 2000US-0218290.	
PR	PR	26-JUL-2000; 2000US-0220963.	
PR	PR	26-JUL-2000; 2000US-0220964.	
PR	PR	14-AUG-2000; 2000US-0224518.	
PR	PR	14-AUG-2000; 2000US-0224519.	
PR	PR	14-AUG-2000; 2000US-0225213.	
PR	PR	14-AUG-2000; 2000US-0225214.	
PR	PR	14-AUG-2000; 2000US-0225266.	
PR	PR	14-AUG-2000; 2000US-0225267.	
PR	PR	14-AUG-2000; 2000US-0225268.	
PR	PR	14-AUG-2000; 2000US-0225270.	
PR	PR	14-AUG-2000; 2000US-0225447.	
PR	PR	14-AUG-2000; 2000US-0225757.	
PR	PR	14-AUG-2000; 2000US-0225758.	
PR	PR	14-AUG-2000; 2000US-0225759.	
PR	PR	18-AUG-2000; 2000US-0226279.	
PR	PR	22-AUG-2000; 2000US-0226681.	
PR	PR	22-AUG-2000; 2000US-0226868.	
PR	PR	22-AUG-2000; 2000US-0227182.	
PR	PR	23-AUG-2000; 2000US-0227009.	
PR	PR	30-AUG-2000; 2000US-0228924.	
PR	PR	01-SEP-2000; 2000US-0229287.	
PR	PR	01-SEP-2000; 2000US-0229343.	
PR	PR	01-SEP-2000; 2000US-0229344.	

PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 10204; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 other;
Query Match 10.5%; Score 295.2; DB 22; Length 13862;
Best Local Similarity 74.6%; Pred. No. 3.7e-40;
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;
QY 1667 AACAGAAATGATAAGCTGCGAACCGTGGCACACGCTCATAGTCTTAGCTGGGAGG 1726
DB 1508 AAACCTTAATAAGCTGGCTCTGTGGCTCATGCTGTGATCCAGCACTTTGGGAGG 1567
QY 1727 TTGAGGAGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCTGGGCAACATAACAG 1786
DB 1568 CTGAGGAGGAGGATGACTTGAGGTCAGGAGTTTGAGACGAGCTGGCCCAACATGGCGAA 1627
QY 1787 ATCCTGTCTCTCAAAAAAAGAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
DB 1628 ACCCATCTCTACTAAAAATACAAAAAGTTAGCCAGGATGGGCGGCGATGGTGGC 1687
QY 1846 TCACGCCCTGTAATCCAGCACTTTGGGAGGCGGAGCGGCGGATCACCTGTGTCAGGA 1905
DB 1688 TCATGCTGTAATCCAGCACTTTGGGAGGCGGAGGTTGGGCGGATCACCTGAGTCA 1747
QY 1906 GTTTGAGACGAGCTGGCCCAACATGCAAAACCCCTCTCTACTCAAAATGCAAAATTA 1965
DB 1748 GTTCAAGACGAGCTGGCAACATGGCAAAACCCCTCTCTCTATATAAAATACAAAATTA 1807
QY 1966 GCCAGGCGTGGTAGCAGGACCTGTAAATCCAGCTACTTTGGGAGGCTGAGGAGGAGAA 2025
DB 1808 GCCAGGCTTATGGTAGGCGCTATATAATCCAGCTACTCGGAAGGCTGAGGAGGAGAA 1867
QY 2026 CGCTTGAACCCAGGAGGTGGAGTTGCAGTAGCTAGAGTCGTGGCGCTTGCCTCCAGCC 2085

[illegible]

PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	06-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

	Query Match	10.4%	Score 292.6;	DB 22;	Length 7809;
	Best Local Similarity	77.2%;	Pred. No. 9.6e-40;		
	Matches 385;	Conservative 0;	Mismatches 104;	Indels 10;	Gaps 2;
Qy	1666	AAACAGAAATGATAAGCTGGGAACCGTGGCACAACGCCTCATAGTTCTTAGCTGTCTGGGAG	1725		
Db	1081	AAAAAGTAAAATCAGGCTGGGATGCTGGTCACGGCTATATCCAGCACACTTTGGGAG	1140		
Qy	1726	GTTGAGGAGGAGGATGGCTTTAACACAGCAGGTGTTCAAGGCCAGCGCTGGGCAACATACAA	1785		
Db	1141	GCTGAGGCAGGAGATCACTTTGATGCCCAGGAGGCTCAAGAGCCAACCTGCACAAAGATGTA	1200		

	Query Match	10.4%	Score 292.6;	DB 22;	Length 7809;
	Best Local Similarity	77.2%;	Pred. No. 9.6e-40;		
	Matches 385;	Conservative 0;	Mismatches 104;	Indels 10;	Gaps 2;
Qy	1666	AAACAGAAATGATAAGCTGGGAACCGTGGCACACGCCTCATAGTTCTTAGCTGTCTGGGAG	1725		
Db	1081	AAAAAGTAAAATCAGGCTGGGATGCTGGTCACGGCTATATCCAGCACACTTTGGGAG	1140		
Qy	1726	GTTGAGGAGGAGGATGGCTTTGAACACAGAGTGTTCAAGGCCAGCGCTGGGCAACAATACAA	1785		
Db	1141	GCTGAGGCAGGAGATCACTTTGATGCCCAGGAGGCTCAAGACCAACTGCACAAAGATGTA	1200		

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX
XX Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 other;

Query Match 10.4%; Score 292.4; DB 22; Length 30393;
Best Local Similarity 78.5%; Pred. No. 1.1e-39;
Matches 377; Conservative 0; Mismatches 96; Indels 7; Gaps 2;
QY 1692 GTGGCACAGCTCATAGTTCTAGCTGTTGGGAGGTTGAGGAGGAGGAGGTTGAGCA 1751
DB 13611 GTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCTGAGGCGGTAATCATTTGAGGT 13552
QY 1752 CAGGTGTTCAAGCCAGCTGGGCAACATAACAAGATCCCTGCTCTCAAAAAA 1811
DB 13551 CAGGAGTTCAAGACAGCTGGCCACATGTTGAACCCCGCTCTCT-----ACTAAAA 13498
QY 1812 AAAAAAGAAAGAGAGAGGCGCGGCTGCTGCTCACGCTCTGTAATCCAGCACTTTGG 1871
DB 13497 TACAAAAGAAAGCTAGGTAGGCGGCGCATGCTGCTCACGCTCTGTAATCCAGCACTTTGG 13438
QY 1872 GAGGCGAGCGGCGGCGATCAGCTGCTGCTCAGAGTTTGAGACAGCCTGGCCACATGG 1931
DB 13437 GAGGCGAGCGGCGGCGATCAATTTGAGGTCAGGAGTTTGAGACAGCCTGGCCACATGG 13378
QY 1932 CAAACCCCGCTGCTACTCAAAATGCAAAATTAGCCAGCGCTGGTAGGAGGACCTGTA 1991
DB 13377 TAAACCTCGTCTCTACTAAACACAAAAATTAGCCAGCGCTGGTAGGAGGACCTGTA 13318
QY 1992 ATCCAGCTACTTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTTGAGGTTG 2051
DB 13317 ATCCAGCTACTCAGGAGGCTGAGCAGGAGTAATGCTTGAATCCAGGCGGCGAAGTTG 13258
QY 2052 CAGTAAGCTGAGATCGTGGCTGCTCAGCTCCAGCCTGGGCGACAGCAAGACTCTCTCT 2111
DB 13257 CAGTGAGCTGAGATCATGCCACTGCACCTCCAGCCTGGGCAAC-AGACAAAGACTCAGTCT 13199
QY 2112 CAGAAAAA 2171
DB 13198 CAAAAA 13139
RESULT 15
ABAI5431
ID ABAI5431 standard; DNA; 25012 BP.
XX AC ABAI5431;
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 7762.
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW anti allergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX Homo sapiens.
XX OS
XX PN WO200159063-A2.
XX PD 16-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01334.
XX
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.

CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 25012 BP; 6932 A; 4835 C; 5560 G; 7685 T; 0 other;

Query Match 10.3%; Score 291; DB 22; Length 25012;

Best Local Similarity 79.6%; Pred. No. 1.9e-39;

Matches 356; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

QY 1681 AGCTCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGGTTGAGAGGGAGGA 1740

Db ||||| | | ||||| || | | | | | ||||| || ||||| || ||||| || |||||

Db 19050 AGCTGGCAGGGTGGCTCATGCTCTAATCCCAACACTGTGGGAGGCCAAGCGGGAGGA 19109

QY 1741 TGGCTTGACACAGTGTTCAGGGCCAGCGCTGGGCAACATAACAAGATCCTGCTCT-CA 1799

Db || | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 19110 TTGCTGAAGTCCAGGAGTTCAGACACAGCGCTGGGAGCATAGCCAGCCCGCTCTATT 19169

QY 1800 AAAAAAATAAAAAAGAGAGAGCGCGGCTGGCTCACGCTGTAATC 1859

Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 19170 TTTTAATTAATAAAAAAATAATGTAGGCCCGGCGACGGTGGCTCACGCTGTAATC 19229

QY 1860 CCAGCACTTTGGAGCGCGAGCGCGGATCACCTGTGTGTCAGGAGTTTGAGACGAGCC 1919

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 19230 CTAGCACTTTGGAGCGTGGAGCGGTGGATCACCGAGTTCAGGAGTTCAAGACGAGCC 19289

QY 1920 TGGCCAAACATGGCAAAACCCGCTCTGTACTCAAAATGCAAAATTTAGCCAGCGGTGGTAG 1979

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 19290 TGGCCAAACATGGTGAACCCCTGTCTCTACTAAAAATACAAAAATTTAGCTGGCGGTGGTGG 19349

QY 1980 CAGGCACCTGTAAATCCAGCTACTTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGG 2039

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 19350 CAGGCGCTGTAAATTCAGCTACTTTGGGAGGCTGAGGAGGAGAAATAGCTTGAACCCAGG 19409

QY 2040 AGGTGGAGGTTGACAGTAAGCTGAGATCGTCCGCTTGCACTCCAGCTGGGCGCAAGAGC 2099

Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 19410 AGGCAGAGGTTATGCTGAGCCGAGATTTGCCATTGCACTCCAGCTGGGCAACAAGAGC 19469

QY 2100 AAGACTCTGCTCAGAAAAAATAAAA 2126

Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 19470 GAAACGCCATCTCCAAAAAATAAGATA 19496

Search completed: June 19, 2003, 00:37:09
Job time : 580.918 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:34:44 ; Search time 3799.83 Seconds

(without alignments)

12010.756 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 2818

Sequence: 1 agagaccatcatattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthma:**

3: em_estlin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_hc:**

9: gb_estli:**

10: gb_est2:**

11: gb_hc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	850.6	30.2	856	9 AL559877	AL559877
2	847	30.1	872	12 BG758510	BG758510 602712721
3	809.4	28.7	953	14 BQ065440	BQ065440 AGENCOURT
4	799.2	28.4	1052	14 BQ055935	BQ055935 AGENCOURT
5	739.4	26.2	743	12 BG686133	BG686133 602638412
6	705.6	25.0	820	12 BG757089	BG757089 602715124

7	705.2	25.0	820	12	BG755526
8	681	24.2	693	12	BF975096
9	667.4	23.7	942	12	BF975166
10	663	23.5	664	12	BG754140
11	655.4	23.3	693	12	BG757292
12	627.6	22.3	670	12	BG341546
13	626	22.2	956	12	BF664355
c 14	572.6	20.3	604	10	AW978582
15	568	20.2	928	12	BG398364
16	562	19.9	570	10	AW504807
17	551.4	19.6	843	12	BG758815
18	546.8	19.4	1034	12	BG755005
19	540	19.2	541	12	BF238155
20	493.6	17.5	511	12	BG170756
21	471	16.7	517	12	BG170824
c 22	470.2	16.7	507	9	AL581406
c 23	467.8	16.6	528	9	AA954956
24	467.6	16.6	939	12	BG341819
25	456.4	16.2	889	12	BG686876
26	453.8	16.1	1140	12	BF664352
c 27	446.2	15.8	476	10	AW135547
c 28	442	15.7	442	9	AT016902
c 29	441.4	15.7	464	10	AW452648
30	411.4	14.6	757	9	AJ446140
31	409.8	14.5	696	9	AJ453647
c 32	405.2	14.4	422	10	AW263139
33	405.4	14.4	522	12	BG144705
34	369.4	13.1	374	10	AW401901
35	366.2	13.0	729	9	AJ450296
36	365.6	13.0	688	9	AJ450317
c 37	358.2	12.7	363	9	AI249134
38	353.4	12.5	413	10	AW768364
c 39	353	12.5	353	9	AA831307
40	352	12.5	650	9	AJ449745
c 41	330.4	11.7	332	9	AA504653
c 42	321	11.4	321	9	AI468242
c 43	317.8	11.3	364	10	AW182255
c 44	308.6	11.0	653	10	BB637360
45	291.8	10.4	623	9	AJ450295

ALIGNMENTS

RESULT 1
AL559877
LOCUS
DEFINITION
AL559877 LTI_FLO11_BCl Homo sapiens cDNA clone CS0DG003YB14 5 prime
856 bp mRNA linear EST 16-FEB-2001
ACCESSION
AL559877
VERSION
AL559877.1 GI:12905793
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 856)
AUTHORS
Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
source
1..856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG003YB14"
/clone_lib="LTI_FLO11_BCl"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"

/note="Vector: pCMVSPORT 6: 1st strand cDNA was primed with a NotI-oligo(dN) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 209 a 217 c 202 g 226 t 2 others

Query Match 30.2%; Score 850.6; DB 9; Length 856;
Best Local Similarity 99.6%; Pred. No. 2.3e-90;
Matches 850; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 62 GACACTCGGACACCACTATGGACAGCCTCTTGATGAACCGGAGAGATTCTTTACCAA 121
DB 1 GACACTCGGACACCACTATGGACAGCCTCTTGATGAACCGGAGAGATTCTTTACCAA 60
QY 122 TTCAAAATGTCGCTGGCTAGGCTGCGGCTGAGACCTACCTGTCTAGTGAAG 181
DB 61 TTCAAAATGTCGCTGGCTAGGCTGCGGCTGAGACCTACCTGTCTAGTGAAG 120
QY 182 AGGCGTGACAGTGCTACATCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACCGC 241
DB 121 AGGCGTGACAGTGCTACATCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACCGC 180
QY 242 TGGCAGCTGGAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTCGGCCG 301
DB 181 TGGCAGCTGGAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTCGGCCG 240
QY 302 TGCTACCGCTCACCCTGGTTCACTCTCTGGAGCCCTGCTACGACTGTCGCGACATGTG 361
DB 241 TGCTACCGCTCACCCTGGTTCACTCTCTGGAGCCCTGCTACGACTGTCGCGACATGTG 300
QY 362 GCCGACTTCTGCGAGGGAACCCCACTCAGTCTGAGGATCTTCACCGCGCGCTCTAC 421
DB 301 GCCGACTTCTGCGAGGGAACCCCACTCAGTCTGAGGATCTTCACCGCGCGCTCTAC 360
QY 422 TTCTGTGAGGACCGCAAGGCTGAGCCCGGAGGCTGCGGCGCTGACCGCGCGGGTG 481
DB 361 TTCTGTGAGGACCGCAAGGCTGAGCCCGGAGGCTGCGGCGCTGACCGCGCGGGTG 420
QY 482 CAAATAGCCATATGACCTTCAAGATATTTTACTGCTGGAATACTTTGTAGAAAAC 541
DB 421 CAAATAGCCATATGACCTTCAAGATATTTTACTGCTGGAATACTTTGTAGAAAAC 480
QY 542 CATGAAGAAGCTTTCAAAGCTGCGAAGGCTGCATGAATTCAGTTCGCTCTCCAGA 601
DB 481 CATGAAGAAGCTTTCAAAGCTGCGAAGGCTGCATGAATTCAGTTCGCTCTCCAGA 540
QY 602 CAGCTTCGGGCGATCCTTTTGGCCCTGTATGAGGTTGATGACTTACGACGCAATTCGT 661
DB 541 CAGCTTCGGGCGATCCTTTTGGCCCTGTATGAGGTTGATGACTTACGACGCAATTCGT 600
QY 662 ACTTTGGGACTTTGATAGCAACTTCAGGAATGTCACACGATGAATATCTCTGCTGA 721
DB 601 ACTTTGGGACTTTGATAGCAACTTCAGGAATGTCACACGATGAATATCTCTGCTGA 660
QY 722 AGACAGTGGTAAAAACAGTCTCTCAAGTCTTCTCTGTTTATTTCTCAACTCTCACT 781
DB 661 AGACAGTGGTAAAAAAGTCTCTCAAGTCTTCTCTGTTTATTTCTCAACTCTCACT 720
QY 782 TTCCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTGA 841
DB 721 TTCCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTGA 780
QY 842 AAATAGAGAGNACACAGGCTGCGCCAGGAGCTGCTCAATTTGGTCAGTTTGAATG 901
DB 781 AAATAGAGAGNACACAGGCTGCGCCAGGAGCTGCTCAATTTGGTCAGTTTGAATG 840
QY 902 CAACATTTGCCCG 914

|||||

841 CAACATTGTCCCY 853

RESULT 2

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 872)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@rs-research.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM1698 row: 1 column: 06

High quality sequence stop: 836.

Location/Qualifiers

1..872

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4853069"

/tissue_type="primary B-cells from tonsils (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pCMV7; Site: 1: XhoI;

Site: 2: EcoRI; CDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-CDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: This is a NIH-MGC Library."

BASE COUNT 211 a 221 c 212 g 228 t

ORIGIN

Query Match 30.1%; Score 847; DB 12; Length 872;

Best Local Similarity 99.3%; Pred. No. 5.9e-90;

Matches 861; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 16 ATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGACACTCTGGACAC 75

DB 2 ATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGACACTCTGGACAC 61

QY 76 CACTATGGACAGCCTCTTGATGAACCGGAGGAGATTTCTTTACCAATTTCAAAATGTCCG 135

DB 62 CACTATGGACAGCCTCTTGATGAACCGGAGGAGATTTCTTTACCAATTTCAAAATGTCCG 121

QY 136 CTGGGCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGCGTCACAGTGC 195

DB 122 CTGGGCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGCGTCACAGTGC 181

QY 196 TACATCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGACGCGTGCACGTGGAAT 255

DB 182 TACATCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGACGCGTGCACGTGGAAT 241

QY 256 GCTCTCTCTCCGCTACATCTCGGACTTGGACCTTAGACCTTGGCCGCTGCTACCGCGTAC 315

Db 242 GCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCCGCTGCTACCGGTCAC 301
QY 316 CTGGTTACCTCCTCGAGCCCTGCTAGACTGTGCCGACATGTGGCGACTTTCGG 375
Db 302 CTGGTTACCTCCTCGAGCCCTGCTAGACTGTGCCGACATGTGGCGACTTTCGG 361
QY 376 AGGGAACCCCAACCTCAGTCTGAGGACTTTCACCCGGCGCTCTACTTCTGTGAGACCG 435
Db 362 AGGGAACCCCAACCTCAGTCTGAGGACTTTCACCCGGCGCTCTACTTCTGTGAGACCG 421
QY 436 CAAGGCTGAGCCCGAGGGCTCGCGGCTGACCGCGCGGGTGCAAAATAGCCATCAT 495
Db 422 CAAGGCTGAGCCCGAGGGCTCGCGGCTGACCGCGCGGGTGCAAAATAGCCATCAT 481
QY 496 GACCTTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTT 555
Db 482 GACCTTCAAGATATATTTTACTGCTGGATACTTTTGTAGAAAACCATGAAGAATTT 541
QY 556 CAAGGCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGTTCGGCGCAT 615
Db 542 CAAGGCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGTTCGGCGCAT 601
QY 616 CTTTTTGGCCCTGTATGAGGTGATGACTTACGACAGCCATTTCTGACTTTGGGACTTTG 675
Db 602 CTTTTTGGCCCTGTATGAGGTGATGACTTACGACAGCCATTTCTGACTTTGGGACTTTG 661
QY 676 ATAGCAACTTCCAGGAATGTACACACAGATGAAATATCTCTGCTGAAGACAGTGGATAA 735
Db 662 ATAGCAACTTCCAGGAATGTACACACAGATGAAATATCTCTGCTGAAGACAGTGGATAA 721
QY 736 AAACAGTCTTCAAGTCTTCTCTGTTTATTTTCACTCTCACTCTCTCTTAGAGTTTAC 795
Db 722 AAACAGTCTTCAAGTCTTCTCTGTTTATTTTCACTCTCACTCTCTCTTAGAGTTTAC 781
QY 796 AGAAAAATATTTATATAGGACTCTTAAAGATCTATGCTTGAAGATAGAGAGAA 855
Db 782 AGAAAAATATTTATATAGGACTCTTAAAGATCTATGCTTGAAGATAGAGAGAA 841
QY 856 CACAGGTCTGGCGAGGAGCTGCTGCA 882
Db 842 CCCCCTGCTGG-CAGGAGCTGCTGCA 867

RESULT 3
LOCUS BQ065440
DEFINITION AGENCOURT_6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
5', mRNA sequence.
ACCESSION BQ065440
VERSION BQ065440.1 GI:19894486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 953)
AUTHORS NIH-MGC <http://mgi.nhl.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLC2108 row: P column: 10
High quality sequence scops: 634.
Location/Qualifiers
1..953
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone_image="5929977"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally cloned
into EORI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 238 a 236 c 233 g 246 t
ORIGIN
Query Match 28.7%; Score 809.4; DB 14; Length 953;
Best Local Similarity 98.5%; Pred. No. 1.3e-85;
Matches 828; Conservative 0; Mismatches 11; Indels 2; Gaps 1;
QY 17 TTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGACACTCTGGACACC 76
Db 1 TTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGACACTCTGGACACC 60
QY 77 ACTATGGACAGCCTCTTGATGAACCGAGGAAAGTTTCTTACCAATTCAAAAATGTCGCG 136
Db 61 ACTATGGACAGCCTCTTGATGAACCGAGGAAAGTTTCTTACCAATTCAAAAATGTCGCG 120
QY 137 TGGGCTAAGGGTGGCGGTGAGACCTACCTGCTAGTGTAGTGAAGAGGGGTGACAGTGTG 196
Db 121 TGGGCTAAGGGTGGCGGTGAGACCTACCTGCTAGTGTAGTGAAGAGGGGTGACAGTGTG 180
QY 197 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGACGGCTGCCACGTGGAAATG 256
Db 181 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGACGGCTGCCACGTGGAAATG 240
QY 257 CTCCTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCCGCTCACC 316
Db 241 CTCCTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCCGCTCACC 300
QY 317 TGGTTCACTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGGACTTTCGCGA 376
Db 301 TGGTTCACTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGGACTTTCGCGA 360
QY 377 GGGAAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGC 436
Db 361 GGGAAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGC 420
QY 437 AAGGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATG 496
Db 421 AAGGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATG 480
QY 497 ACCTTCAAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTC 556
Db 481 ACCTTCAAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTC 540
QY 557 AAAGCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATC 616
Db 541 AAAGCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATC 600
QY 617 CTTTTCGCCCTGTATGAGGTGTATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 676
Db 601 CTTTTCGCCCTGTATGAGGTGTATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 660
QY 677 TAGCAACTTCCAGGAATGTACACACAGATGAAATATCTCTGCTGAAGACAGTGGATAAAA 736
Db 661 TAGCAACTTCCAGGAATGTACACACAGATGAAATATCTCTGCTGAAGACAGTGGATAAAA 720
QY 737 ACAGCTCTTCAAGTCTTCTCTGTTTATTTTCACTCTCACTTCTTAGAGTTTACA 796
Db 721 AACAGTCTTCAAGTCTTCTCTGTTTATTTTCACTCTCACTTCTTAGAGTTTAC 780


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT      176 a   197 c   188 g   182 t
ORIGIN

Query Match      26.2%; Score 739.4; DB 12; Length 743;
Best Local Similarity 99.9%; Pred. No. 2.1e-77;
Matches 740; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  4 GAACCATCATTAATGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGA 63
Db  2 GAACCATCATTAATGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGA 61
QY  64 CACTCTGGACACCATATGACAGCCCTTTGATGAACCGAGGAGGAACTTTCTTACCAATT 123
Db  62 CACTCTGGACACCATATGACAGCCCTTTGATGAACCGAGGAGGAACTTTCTTACCAATT 121
QY  124 CAAAATATGTCGTGGCTAAGGCTGGCGGTGAGACCTACCTGTGCTAGCTAGTGAAGAG 183
Db  122 CAAAATATGTCGTGGCTAAGGCTGGCGGTGAGACCTACCTGTGCTAGCTAGTGAAGAG 181
QY  184 GCGTGACAGTGTACATCTCTTTCACTGGACCTTTGTTATCTTCGCAATAAAGAACGGCTG 243
Db  182 GCGTGACAGTGTACATCTCTTTCACTGGACCTTTGTTATCTTCGCAATAAAGAACGGCTG 241
QY  244 CCACGTGGAAATGCTCTTCCCTCGCTPACATCTCGGACTGGGACTGAGACCTGAGCCGTGGCGGCTG 303
Db  242 CCACGTGGAAATGCTCTTCCCTCGCTPACATCTCGGACTGGGACTGAGACCTGAGCCGTGGCGGCTG 301
QY  304 CTACCGGCTACCTGGTTCCACCTCTCTGGAGCCCTCTCTACGACTGTGCCCGACATGTGGC 363
Db  302 CTACCGGCTACCTGGTTCCACCTCTCTGGAGCCCTCTCTACGACTGTGCCCGACATGTGGC 361
QY  364 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTT 423
Db  362 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTT 421
QY  424 CTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGCGCGGCTGTCACCGCGCGGGGTGCA 483
Db  422 CTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGCGCGGCTGTCACCGCGCGGGGTGCA 481
QY  484 AATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATFACTTTTGTAGAAAACCA 543
Db  482 AATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATFACTTTTGTAGAAAACCA 541
QY  544 TGAAGAAGCTTTCAAGCCCTGGGAAGGCTGCATGAAAATTCAGTTCTCTCCAGACA 603
Db  542 TGAAGAAGCTTTCAAGCCCTGGGAAGGCTGCATGAAAATTCAGTTCTCTCTCCAGACA 601
QY  604 GCTTCGGCGCATCTCTTTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTAC 663
Db  602 GCTTCGGCGCATCTCTTTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTAC 661
QY  664 TTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTGAAG 723
Db  662 TTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTGAAG 721
QY  724 ACAGTGGATAAAAAACAGTCC 744
Db  722 ACAGTGGATAAAAAACAGTCC 742
```

```
RESULT 6
BG757089
LOCUS      820 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602715124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
mRNA sequence.
ACCESSION  BG757089
VERSION     BG757089.1 GI:14067742
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 820)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Ph.D.
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1704 row: o column: 06
            High quality sequence stop: 675.
```

FEATURES

```
source
1..820
    Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4855517"
        /clone_lib="NIH_MGC_48"
        /tissue_type="primary B-cells from tonsils (cell line)"
        /lab_host="DH10B (phage-resistant)"
        /notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
        Site:2: EcoRI; cDNA made by oligo-dt priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCACGAG(G). Size-selected >500bp
        for average insert size 1.8kb. Library constructed by Ling
        Hong in the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies).
        Note: this is a NIH_MGC Library."
BASE COUNT      210 a   230 c   200 g   180 t
```

```
Query Match      25.0%; Score 705.6; DB 12; Length 820;
Best Local Similarity 96.5%; Pred. No. 1.7e-73;
Matches 753; Conservative 0; Mismatches 24; Indels 3; Gaps 3;

QY  1 AGAAGACCATCATTAATGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGA 60
Db  6 AGAAGACCATCATTAATGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGA 65
QY  61 AGACACTCTGGACACCATATGACAGCCCTTTCATGAACCGGAGGAGTTCTTTACCA 120
Db  66 AGACACTCTGGACACCATATGACAGCCCTTTCATGAACCGGAGGAGTTCTTTACCA 125
QY  121 ATTCAAAAATGTCGGCTGGGCTAAGGCTGCGCTGAGACCTTACCTGTGCTAGTAGTAA 180
Db  126 ATTCAAAAATGTCGGCTGGGCTAAGGCTGCGCTGAGACCTTACCTGTGCTAGTAGTAA 185
QY  181 GAGGCGGTGACAGTGTACATCTTTTTCACATGGAGCTTTGGTGTATCTTCGCAATAAAGAGG 240
Db  186 GAGGCGGTGACAGTGTACATCTTTTTCACATGGAGCTTTGGTGTATCTTCGCAATAAAGAGG 245
QY  241 CTGCCAGCTGGAATGCTCTCTCCCTACATCTCGGACTGGGACCTAGACCTTGCGCG 300
Db  246 CTGCCAGCTGGAATGCTCTCTCCCTACATCTCGGACTGGGACCTAGACCTTGCGCG 305
```

```
QY 301 CTGCTACCGCGGTCACTGGTTCACTCTCTGGAGCCCTGTACGACTGTGCCGACATGT 360
Db 306 CTGCTACCGCGGTCACTGGTTCACTCTCTGGAGCCCTGTACGACTGTGCCGACATGT 365
QY 361 GGCCGACTTTCTGGAGGGAACCCCAACCTCACTCTGAGATCTTACCGCGGCTCTTA 420
Db 366 GGCCGACTTTCTGGAGGGAACCCCAACCTCACTCTGAGATCTTACCGCGGCTCTTA 425
QY 421 CTTCTGTGAGACCGCAAGGCTGAGCCGAGGGCTGCGGCGCTGCACCGCGCGGGT 480
Db 426 CTTCTGTGAGACCGCAAGGCTGAGCCGAGGGCTGCGGCGCTGCACCGCGCGGGT 485
QY 481 GCAATAGCCATCATGACCTTCAAGATATATTTTACTGCTGGAATCTTTGTAGAAA 540
Db 486 GCAATAGCCATCATGACCTTCAAGATATATTTTACTGCTGGAATCTTTGTAGAAA 545
QY 541 CCATGAAGAACTTCAAGCCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAG 600
Db 546 CCATGAAGAACTTCAAGCCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAG 605
QY 601 ACAGCTTGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCAATTGC 660
Db 606 ACAGCTTGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCA-TTCG 664
QY 661 TACTTTGGGACTTTGTATAGCACTTCCAGGAATGTACACA-CGATGAATATCTCTGCT 719
Db 665 TACTTTAGGACTTTCGATAGCAAACTCCAGGAATGTACACAGCGATGAATATCTCTGCT 724
QY 720 GAAGACAGTGGATAA-AAACACTCTCTCAAGCTTCTCTGTTTATTTCTTCACTCTC 778
Db 725 GAAGACAGGCGATAACAAACGACCTCAAAGTCAACACTGAAGACAGCGCGTCAACGCC 784
```

RESULT 7

BG755526

LOCUS

DEFINITION

602716206F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4856568 5',

mRNA sequence.

BG755526

BG755526.1 GI:14066179

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 820)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: L1CM1707 row: k column: 01

High quality sequence stop: 638.

Location/Qualifiers

1..820

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4856568"

/clone_lib="NIH_MGC_48"

/tissue_type="primary B-cells from tonsils (cell line)"

/lab_host="PH10B (phage-resistant)"

/notes="Organ: B-cells; Vector: pORF7; Site 1: xhoi;

Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

FEATURES

source

RESULT 8

BF975096

LOCUS

DEFINITION

BF975096

60245679F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:436722 5',

mRNA sequence.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

693 bp

mRNA

linear

EST 22-JAN-2001

IMAGE:436722 5',

mRNA sequence.

Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library.

BASE COUNT 245 a 161 c 168 g 246 t

Query Match 25.0%; Score 705.2; DB 12; Length 820;

Best Local Similarity 94.5%; Pred. No. 1.9e-73;

Matches 775; Conservative 0; Mismatches 38; Indels 7; Gaps 4;

QY 634 GGTGTGACTTACGAGACGATTTTCGTACTTTGGGACTTTTGATAGCAACTTCCAGGAAT 693

Db 2 GGTGTGACTTACGAGACGATTTTCGTACTTTGGGACTTTTGATAGCAACTTCCAGGAAT 60

QY 694 GTCACACAGCATGAATATCTCTGCTGAACACAGTGGATAAACAACAGTCTCTCAAGTCT 753

Db 61 GTCACACAGCATGAATATCTCTGCTGAACACAGTGGATAAACAACAGTCTCTCAAGTCT 120

QY 754 TCTCTGTTTTTATTCTTCAACTCTCACTTTCTAGAGTTTACAGAAAAAATATTATATA 813

Db 121 TCTCTGTTTTTATTCTTCAACTCTCACTTTCTAGAGTTTACAGAAAAAATATTATATA 180

QY 814 CGACTCTTTAAAAAGATCTATGCTTTGAAAATAGAGAAGAACACAGGCTCGGCCAGGA 873

Db 181 CGACTCTTTAAAAAGATCTATGCTTTGAAAATAGAGAAGAACACAGGCTCGGCCAGGA 240

QY 874 CGTCTGCAATTCGTGCGAGTTTGAATGCAACATGTCCTCTACTGGGAATAACAGAACT 933

Db 241 CGTCTGCAATTCGTGCGAGTTTGAATGCAACATGTCCTCTACTGGGAATAACAGAACT 300

QY 934 GCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAG 993

Db 301 GCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAG 360

QY 994 AGCAGAAGGTAGATCTCTAAAGACATGGTGAGAGGATCAAAATGTTTTTATATCAACATCC 1053

Db 361 AGCAGAAGGTAGATCTCTAAAGACATGGTGAGAGGATCAAAATGTTTTTATATCAACATCC 420

QY 1054 TTTTATTATTGATTTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATCTCTTTT 1113

Db 421 TTTTATTATTGATTTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATCTCTTTT 480

QY 1114 CCCTTGACGTTTACTTTTCAAGTAACACAACTCTCCATCAGGCCATGATCTATAGACC 1173

Db 481 CCCTTGACGTTTACTTTTCAAGTAACACAACTCTCCATCAGGCCATGATCTATAGACC 540

QY 1174 TCCTAATGAGAGTATCTGGGTGATTTGTACCCCAACCATCTCTCCAAAGCATTAATATC 1233

Db 541 TCCTAATGAGAGTATCTGGGTGATTTGTACCCCAACCATCTCTCCAAAGCATTAATATC 600

QY 1234 CAATCATGCGCTGATGTTTAAATACGAGAAGCATGTTTTTATGTTGTACAAAAGAG 1293

Db 601 CAATCATGCGCTGATGTTTAAATACGAGAAGCATGTTTTTATGTTGTACAAAAGAG 660

QY 1294 ATTGTATGGGTGGGATGGAGGTAT---AGACCATGCGATGGTCACTCTCAAG--CTACT 1348

Db 661 ATTGTATGGGTGGGATGGGCGCTCTCAGACCATGCGATGGTCACTCTCAAGGCTACTTT 720

QY 1349 TTAATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAAAGACCCCTTAATATGGGTT 1408

Db 721 TTAATAAAGGATCTTCAATGGGAGGAGGACTGTGAACAAAGGAAATCTTAAATAATGGGT 780

QY 1409 GATGTCCTGAAG-TAGCAAAATCTTCTGGAACGCAAACTCT 1447

Db 781 GATGTCCTGAAGTTAGCAAAATCTTCTGGAACGCAAACTCT 820


```
ACCESSION      BF975096
VERSION        BF975096.1  GI:12342311
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 693)
JOURNAL        NIH-MGC http://mgi.nci.nih.gov/
COMMENT        National Institutes of Health, Mammalian Gene Collection (MGC)
                Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgabbs-remail.nih.gov
                Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLCMI209 row: n column: 19
                High quality sequence stop: 692.
                Location/Qualifiers
                1..693
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4336722"
                /clone_lib="NIH_MGC_48"
                /tissue_type="primary B-cells from tonsils (cell line)"
                /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dt priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
                179 a 167 c 162 g 185 t
BASE COUNT    179 a 167 c 162 g 185 t
ORIGIN
Query Match      24.2%; Score 681; DB 12; Length 693;
Best Local Similarity 99.9%; Pred. No. 1.4e-70;
Matches 692; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 304 CTACCGCGTACCTGTTTACCTCTCGAGCCCTCTAGACTGTGCCGACATGTGGC 363
DB 1 CTACCGCGTACCTGTTTACCTCTCGAGCCCTCTAGACTGTGCCGACATGTGGC 60
QY 364 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 423
DB 61 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 120
QY 424 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCA 483
DB 121 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCA 180
QY 484 AATAGCCATCATGACCTTCAAGATTTATTTTACTGCTGGAATACTTTGTAGAAACCA 543
DB 181 AATAGCCATCATGACCTTCAAGATTTATTTTACTGCTGGAATACTTTGTAGAAACCA 240
QY 544 TGAAGAAGCTTTCAAAGCTGGGAAGGCTGCGATGAAATTCAGTTCGTCTCCACAGACA 603
DB 241 TGAAGAAGCTTTCAAAGCTGGGAAGGCTGCGATGAAATTCAGTTCGTCTCCACAGACA 300
QY 604 GCTTCGCGCATCTTTTCCCTCTGTATGAGTTGATGACTTACGAGACGCAATTCGTGAC 663
DB 301 GCTTCGCGCATCTTTTCCCTCTGTATGAGTTGATGACTTACGAGACGCAATTCGTGAC 360
QY 664 TTTGGACTTTGATACCACTTCCAGGAATGTACACAGATGAATATCTCGTCGAAG 723
DB 361 TTTGGACTTTGATACCACTTCCAGGAATGTACACAGATGAATATCTCGTCGAAG 420
```

```
QY 724 ACAGTGATATAAAACAGTCTCTCAAGTCTTCTGTTTATTTCTTCAACTCTCCTTT 783
DB 421 ACAGTGATATAAAACAGTCTCTCAAGTCTTCTGTTTATTTCTTCAACTCTCCTTT 480
QY 784 CTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCATATCTTTG 843
DB 481 CTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCATATCTTTG 540
QY 844 ATAGAGAAGAACACAGCTGTGGCCAGGAGCTGCTGCAATTTGGTGCAGTTTGGATCA 903
DB 541 ATAGAGAAGAACACAGCTGTGGCCAGGAGCTGCTGCAATTTGGTGCAG-TTTGAATCA 599
QY 904 ACATTGTCCTCTACTGGGAATAACAGAACTCAGAGCTCGGAGCATCTTAAAGTGTC 963
DB 600 ACATTGTCCTCTACTGGGAATAACAGAACTCAGAGCTCGGAGCATCTTAAAGTGTC 659
QY 964 CGTTTTTCTATGACTTTTAGGTAGGATGAGAGC 996
DB 660 CGTTTTTCTATGACTTTTAGGTAGGATGAGAGC 692
```

```
RESULT 9
LOCUS      BF975166
DEFINITION 602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4335639 5',
            mRNA sequence.
ACCESSION  BF975166
VERSION     BF975166.1  GI:12342381
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 942)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI207 row: a column: 16
            High quality sequence stop: 707.
            Location/Qualifiers
            1..942
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4335639"
            /clone_lib="NIH_MGC_48"
            /tissue_type="primary B-cells from tonsils (cell line)"
            /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
            Site_2: EcoRI; cDNA made by oligo-dt priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Size-selected by Ling
            Hong in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
            249 a 209 c 251 g 233 t
BASE COUNT  249 a 209 c 251 g 233 t
ORIGIN
Query Match      23.7%; Score 667.4; DB 12; Length 942;
Best Local Similarity 91.3%; Pred. No. 4.1e-69;
Matches 799; Conservative 0; Mismatches 61; Indels 15; Gaps 8;
```


Db 482 CCCTTGACGCTTACTTCAAGTAACACAAACACTCTTCCATCAGGCCATGATCTATAGGACC 541

QY 1174 TCCTAATGAGAGTATCTGGTGATTTGAGCCCCAAACCATCTCTCCAAAGCATTAATATC 1233

Db 542 TCCTAATGAGAGTATCTGGTGATTTGAGCCCCAAACCATCTCTCCAAAGCATTAATATC 601

QY 1234 CAATCATGCGCTGATGTTTTAATCAGCAGAGAGCATGTTTTTATGTTTATGTTTACAAAAAGAAG 1293

Db 602 CAATCATGCGCTGATGTTTTAATCAGCAGAGAGCATGTTTTTATGTTTATGTTTACAAAAAGAAG 661

QY 1294 ATT 1296

Db 662 ATT 664

RESULT 11

BG757392

LOCUS 602711022F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851580 5',

DEFINITION mRNA sequence.

ACCESSION BG757392

VERSION BG757392.1 GI:14068045

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 693)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCMI694 row: k column: 05
High quality sequence stop: 693.

FEATURES

source

1. .693

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4851580"

/clone_lib="NIH_MGC_48"

/tissue_type="primary B-cells from tonsils (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 157 a 188 c 178 g 170 t

ORIGIN

Query Match 23.3%; Score 655.4; DB 12; Length 693;

Best Local Similarity 99.7%; Pred. No. 1.3e-67;

Matches 667; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AGAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTTGCAGGGAGGCAAGA 60

Db 2 AGAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTTGCAGGGAGGCAAGA 61

QY 61 AGACACTCTGGACACCACTATGACAGCCCTCTTGATGACCGGAGGAAGTTCTTTACCA 120

Db 62 AGACACTCTGGACACCACTATGACAGCCCTCTTGATGACCGGAGGAAGTTCTTTACCA 121

QY 121 ATTCAAAAATGTCGCTGGGCTAAGGCTCGCGTGGAGACCTACCTGTGCTACGTAGTGAA 180

Db 122 ATTCAAAAATGTCGCTGGGCTAAGGCTCGCGTGGAGACCTACCTGTGCTACGTAGTGAA 181

QY 181 -GAGGCGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGC 239

Db 182 CGAGGGGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGC 241

QY 240 GCTGCCACGTGGAAATGCTCTTCCTCGCTACATCTCGAGCTGGAGCTAGACCTGACCC 299

Db 242 GCTGCCACGTGGAAATGCTCTTCCTCGCTACATCTCGAGCTGGAGCTAGACCTGACCC 301

QY 300 GCTGCTACCCGCTCACCTGGTTTCACTTCCTGGAGCCCTGCTACGACTGTGCCGACATG 359

Db 302 GCTGCTACCCGCTCACCTGGTTTCACTTCCTGGAGCCCTGCTACGACTGTGCCGACATG 361

QY 360 TGGCCGACTTTTTCGCGAGGGAACCCCAACCTCAGCTGAGGATCTTCAACCGCGGCTCT 419

Db 362 TGGCCGACTTTTTCGCGAGGGAACCCCAACCTCAGCTGAGGATCTTCAACCGCGGCTCT 421

QY 420 ACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGG 479

Db 422 ACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGG 481

QY 480 TGCAATATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATACTTTTGTAGAAA 539

Db 482 TGCAATATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATACTTTTGTAGAAA 541

QY 540 ACCATGAAGAAGCTTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCTGCTCTCCA 599

Db 542 ACCATGAAGAAGCTTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCTGCTCTCCA 601

QY 600 GACAGCTTCGCGCATCCTTTTGCCTGATGAGTTGATGACTTAGGACGACCATTC 659

Db 602 GACAGCTTCGCGCATCCTTTTGCCTGATGAGTTGATGACTTAGGACGACCATTC 661

QY 660 GTACTTTGG 668

Db 662 GTACTTTGG 670

RESULT 12

BG341546

LOCUS 602463652F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4576548 5',

DEFINITION mRNA sequence.

ACCESSION BG341546

VERSION BG341546.1 GI:13147984

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 670)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCMI289 row: g column: 13
High quality sequence stop: 650.

FEATURES

source

1. .670

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

```

/clone="IMAGE:4576548"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: xhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      200 a 122 c 134 g 214 t
ORIGIN

```

```

Query Match      22.3%; Score 627.6; DB 12; Length 670;
Best Local Similarity 98.8%; Pred. No. 2.3e-64;
Matches 664; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

QY 634 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 693
DB 2 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 60

QY 694 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTTCAAGTCT 753
DB 61 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTTCAAGTCT 120

QY 754 TCTCTGTTTTTATCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATA 813
DB 121 TCTCTGTTTTTATCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATA 180

QY 814 CGACTCTTTAAAAAGATCTATGCTTGAATAATAGAGAAGAACACAGGCTTGCCAGGGA 873
DB 181 CGACTCTTTAAAAAGATCTATGCTTGAATAATAGAGAAGAACACAGGCTTGCCAGGGA 240

QY 874 CGTCTGCAATTCGGTCAGTTTTGTAATGCAACATTTGCCCTACTGGGAATAACAGAACT 933
DB 241 CGTCTGCAATTCGGTCAGTTTTGTAATGCAACATTTGCCCTACTGGGAATAACAGAACT 300

QY 934 GCAGGACCTGGGAGCATCTCTAAAGTCTCAACGTTTTTCTATGACTTTTAGGTAGGATGAG 993
DB 301 GCAGGACCTGGGAGCATCTCTAAAGTCTCAACGTTTTTCTATGACTTTTAGGTAGGATGAG 360

QY 994 AGCAGAAGTGTAGTCTCTAAAAAGCATGGTGAGAGGATCAAAATGTTTTTATATCAACATCC 1053
DB 361 AGCAGAAGTGTAGTCTCTAAAAAGCATGGTGAGAGGATCAAAATGTTTTTATATCAACATCC 420

QY 1054 TTTATTATTGATTCATTTGAGTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1113
DB 421 TTTATTATTGATTCATTTGAGTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 480

QY 1114 CCCTTGACGTTTACTTTCAAGTAAACAAACTCTTCCATCAGGCCATGATCTATAGACC 1173
DB 481 CCCTTGACGTTTACTTTCAAGTAAACAAACTCTTCCATCAGGCCATGATCTATAGACC 540

QY 1174 TCCTAATAGAGTATCTGGGTGATTTGACCCCAACCATCTCTCC - AAAGCATTAATAT 1232
DB 541 TCCTAATAGAGTATCTGGGTGATTTGACCCCAACCATCTCTCC - AAAGCATTAATAT 600

QY 1233 CCAATCATGGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAGAA 1292
DB 601 CCAATCATGGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAGAA 658

QY 1293 GATTGTTATGGG 1304
DB 659 GATTGTTATGGG 670

```

```

RESULT 13
BF664355
LOCUS
DEFINITION 602146013F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4309707 5',

```

```

mRNA sequence.
BF664355
BF664355.1 GI:11938351
EST.
HOMO
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 956)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM183 row: i column: 04
High quality sequence stop: 725.
FEATURES
Location/Qualifiers
1..956
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4309707"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: xhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 277 a 183 c 219 g 277 t
ORIGIN

```

```

Query Match      22.2%; Score 626; DB 12; Length 956;
Best Local Similarity 97.2%; Pred. No. 2.6e-64;
Matches 690; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

QY 634 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 693
DB 2 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 60

QY 694 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTTCAA - GTC 752
DB 61 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTTCAAAGTCT 120

QY 753 TTTCTGTTTTTATTTCTTCAACTCTCACATTTCTTAGAGTTTACAGAAAAAATTTATAT 812
DB 121 TTTCTGTTTTTATTTCTTCAACTCTCACATTTCTTAGAGTTTACAGAAAAAATTTATAT 180

QY 813 AGCACTCTTTAAAAAGATCTATGCTTTGAAAATAGAGAAGAACACAGGCTTGGCCAGG 872
DB 181 AGCACTCTTTAAAAAGATCTATGCTTTGAAAATAGAGAAGAACACAGGCTTGGCCAGG 240

QY 873 AGTGTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCTCCCTACTGCGAATAACAGAAC 932
DB 241 AGTGTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCTCCCTACTGCGAATAACAGAAC 300

QY 933 TGCAGGACCTGGGAGCATCTCTAAAGTGTCAAGCTTTTTTCTATGACTTTTAGGTAGGATGA 992
DB 301 TGCAGGACCTGGGAGCATCTCTAAAGTGTCAAGCTTTTTTCTATGACTTTTAGGTAGGATGA 360

QY 993 GAGCAGAAGTGTAGTCTCTAAAAAGCATGGTGAGAGCATCAATGTTTTTATATCAACATC 1052
DB 361 GAGCAGAAGTGTAGTCTCTAAAAAGCATGGTGAGAGCATCAATGTTTTTATATCAACATC 1052

```

```

Db      361 GAGCAGAGGTAGATCCTAAAGCATGGTGAGAGATCAAAATGTTTTATATCAACATC 420
QY      1053 CTTTATTATTGATTCATT -TGAGTTAACAGTGGTGTAGTGATAGATTTTCTATTCTT 1111
Db      421 CTTTATTATTGATTCATTGTGAGTTAACAGTGGTGTAGTGATAGATTTTCTATTCTT 480
QY      1112 TTCCCTTGAGCTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGA 1171
Db      481 TTCCCTTGAGCTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGA 540
QY      1172 CCTCCTTAAGAGAGTATCTGGGTGATTTGACCCCAACCATCTCTCCAAAGCATTAATA 1231
Db      541 CCTCCTTAAGAGAGTATCTGGGTGATTTGACCCCAACCATCTCTCCAAAGCATTAATA 599
QY      1232 TCCAAATCATGCGCTGATGTTTAAATCAGCAGAGCATGTTTATGTTTGTGTACAAAGA 1291
Db      600 TCCATCATGCGCTGATGTTTAAATCAGCAGAGCATGTTTATGTTTGTGTACAAAGA 658
QY      1292 AGATTGTTTATGGGTGGGATGGAGGTATAGACCATGATGGTCAACCTTCA 1341
Db      659 AGATGTTATGGGTGGGATGGAGGTATAGACCATGATGGTCAACCTTCA 708

RESULT 14
AW978582/c
LOCUS      AW978582      604 bp      mRNA      linear      EST 02-JUN-2000
DEFINITION      EST390691 MAGE resequences, MAGP Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW978582
VERSION        AW978582.1 GI:8169850
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Hoit
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE        Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL      Unpublished (2000)
COMMENT      Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnqu@tigr.org
Plate: 394
Seq primer: Forward.

FEATURES
source
Location/Qualifiers
1..604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGP"
/note="Vector: pBluescriptSkm"
BASE COUNT      190 a 121 c 109 g 184 t
ORIGIN
Query Match      20.3%; Score 572.6; DB 10; Length 604;
Best Local Similarity 97.6%; Pred. No. 6e-58;
Matches 581; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      1061 TTTGATTATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATTCTTTCCCTTGA 1120
Db      595 TTGAATCCATTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATTCTTTCCCTTGA 536
QY      1121 CTTTACTTTCAAGTAACACAAACTCTCCATCAGGCCATGATCTATAGGACCTCCTAAT 1180
Db      535 CGTTTATTTTCAAGTAACACAAACTCTCCATCAGGCCATGTTTATAGGACCTCCTAAT 476
QY      1181 GAGAGTATCTGGGTGATTTGACCCCAACCATCTCTCCAAAGCATTAATATCAATCAT 1240
Db      475 GAGAGTATCTGGGTGATTTGACCCCAACCATCTCTCCAAAGCATTAATATCAATCAT 416

```

```

QY      1241 GCGCTGATGTTTAAATCAGCAGAGCATGTTTATGTTTGTACAAAAGAGATTGTTA 1300
Db      415 GCGCTGATGTTTAAATCAGCAGAGCATGTTTATGTTTGTACAAAAGAGATTGTTA 356
QY      1301 TGGGTGGGATGGAGGTATAGACCATGATGGTCAACCTTCAAGCTACTTTTAAATAAGGAT 1360
Db      355 TGGGTGGGATGGAGGTATAGACCATGATGGTCAACCTTCAAGCTACTTTTAAATAAGGAT 296
QY      1361 CTTAAAATGGCAGGAGACTGTCAACAAGACACCCCTAAATATGGGTTGATGCTCTGAAGT 1420
Db      295 CTTAAAATGGCAGGAGACTGTCAACAAGACACCCCTAAATATGGGTTGATGCTCTGAAGT 236
QY      1421 AGCAAAATCTTCTGAAAACGCAAACTCTTTTAAAGGAAGTCCCTAAATTTAGAAACACCCACA 1480
Db      235 AGCAAAATCTTCTGAAAACGCAAACTCTTTTAAAGGAAGTCCCTAAATTTAGAAACACCCACA 176
QY      1481 AACTTCACATATCAATAATTAGCAAAACAATTTGGAAGAGAGTTGCTTGAATTTGGGGAGAG 1540
Db      175 AACTTCACATATCAATAATTAGCAAAACAATTTGGAAGAGAGTTGCTTGAATTTGGGGAGAG 116
QY      1541 GAAATATCTATGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTT 1600
Db      115 GAAATATCTATGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTT 56
QY      1601 GCTACATTTTGTATGTTGTCATGCTTCTCCCAAAGGTATATTAACTATATAAGA 1655
Db      55 GCTACATTTTGTATGTTGTCATGCTTCTCCCAAAGGTATATTAACTATATAANTA 1

RESULT 15
BG398364
LOCUS      BG398364      928 bp      mRNA      linear      EST 12-MAR-2001
DEFINITION      602439832F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4566366 5',
mRNA sequence.
ACCESSION      BG398364
VERSION        BG398364.1 GI:13291812
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI284 row: o column: 07
High quality sequence stop: 650.

FEATURES
source
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4566366"
/tissue_type="primary B-cells from tonsils (cell line)"
/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

Note: this is a NIH_MGC Library."

BASE COUNT	280 a	163 c	222 g	263 t
ORIGIN				
Query Match	20.2%	Score 568;	DB 12;	Length 928;
Best Local Similarity	98.6%	Pred. No. 1.5e-57;		
Matches 615;	Conservative 0;	Mismatches 5;	Indels 4;	Gaps 4;
Qy	622	GGCCCTGTATGAGGTTGATGACATTACGAGACGCAATTCGCTACTTTGGGACTTTGATAGCA	681	
Db	55	GGCCCTGTATGAGGTTGATGACATTACGAGACGCAATTCGCTACTTTGGGACTTTGATAGCA	113	
Qy	682	ACTTCCAGGAATGCACACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAG	741	
Db	114	ACTTCCAGGAATGCACACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAG	173	
Qy	742	TCCTTCAAGTCTTCTCTGTTTTTATCTTCAACTCTCAGTTTCTTAGAGTTTACAGAAA	801	
Db	174	TCCTTCAAGTCTTCTCTGTTTTTATCTTCAACTCTCAGTTTCTTAGAGTTTACAGAAA	233	
Qy	802	AATATTATATAGGACTCTTTAAAAAGATCTATGTCTTGAATAATAGAGAGAACACAGG	861	
Db	234	AATATTATATAGGACTCTTTAAAAAGATCTATGTCTTGAATAATAGAGAGAACACAGG	293	
Qy	862	TCGGCCAGGAGCGTCTGCAATTGGTGCAGTTTTTGAATGCAACATTTGCCCTACTGGG	921	
Db	294	TCGGCCCA -GGACGTGCTGCAATTGGTGCAGTTTTTGAATGCAACATTTGCCCTACTGGG	352	
Qy	922	AATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTT	981	
Db	353	AATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTT	412	
Qy	982	AGGTAGGATGAGAGCAGAGGTAGATCCTAAAAAGCATGCTGAGAGGATCAAAATG-TTTT	1040	
Db	413	AGGTAGGATGAGAGCAGAGGTAGATCCTAAAAAGCATGCTGAGAGGATCAAAATGTTTT	472	
Qy	1041	TATATCAACATCCTTTTATTATTGATTCATTTGAGTTAACAGTGGTGTAGTGATAGATT	1100	
Db	473	TATATCAACATCCTTTTATTATTGATTCATTTGAGTTAACAGTGGTGTAGTGATAGATT	532	
Qy	1101	TTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCAT	1160	
Db	533	TTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCAT	592	
Qy	1161	GATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAAACATCTCTCCA	1220	
Db	593	GATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAAACATCTCTCCA	1220	
Qy	1221	AAGCATTAATATCAATCATGCGC	1244	
Db	652	AAGCATTAATATCAATCATGCGC	675	

Search completed: June 19, 2003, 11:07:25
Job time : 3805.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:01:09 ; Search time 118.35 Seconds
(without alignments)
7302.178 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 2818
Sequence: 1 agagaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/ptodata/1/ina/5A-COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B-COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A-COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B-COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PTUS-COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	280	9.9	15297	4	US-09-817-180-3
2	274.2	9.7	11811	4	US-09-078-294-7
3	269.8	9.6	43950	4	US-09-735-934A-3
4	269.2	9.6	99500	4	US-09-798-096-10
5	267.4	9.5	59065	4	US-09-813-817-3
6	267.4	9.5	59065	4	US-09-978-197-3
7	265.6	9.4	9365	4	US-09-608-285A-8
8	265.6	9.4	9365	4	US-09-350-836B-8
9	265.6	9.4	9365	4	US-09-370-265-8
10	265.6	9.4	14747	4	US-09-608-285A-42
11	265.6	9.4	15977	4	US-09-608-285A-59
12	262.2	9.3	21234	4	US-09-810-671-3
13	259.2	9.2	3885	1	US-08-688-145-1
14	258.6	9.2	43950	4	US-09-735-934A-3
15	257.6	9.1	99500	4	US-09-798-096-10
16	257.2	9.1	10684	3	US-08-618-100B-3
17	256.8	9.1	8174	1	US-07-914-281-5
18	256.8	9.1	8174	1	US-08-393-246-5
19	256.8	9.1	8174	1	US-08-525-058A-5
20	256.8	9.1	8174	2	US-08-696-731-5
21	256.8	9.1	8174	4	US-09-042-531-5
22	256.8	9.1	8174	5	PCT-US91-00899-3
23	256.8	9.1	29629	4	US-09-729-995-3
24	254.4	9.0	98844	4	US-09-791-211-10
25	253.8	9.0	16063	4	US-09-801-052-3
26	253.2	9.0	14581	4	US-08-520-373D-4
27	253.2	9.0	22481	4	US-08-367-841A-43

C	28	253.2	9.0	22481	5	PCT-US95-07201-43	Sequence 43, Appl
C	29	253.2	8.9	22484	4	US-09-875-223-2	Sequence 2, Appl
	30	252.2	8.9	7210	2	US-08-257-963B-10	Sequence 10, Appl
	31	252.2	8.9	7210	4	US-08-367-841A-10	Sequence 10, Appl
	32	252.2	8.9	7210	5	PCT-US95-07201-10	Sequence 10, Appl
	33	252.2	8.9	14581	4	US-08-520-373D-4	Sequence 4, Appl
	34	252.2	8.9	22481	4	US-08-367-841A-43	Sequence 43, Appl
	35	252.2	8.9	22481	5	PCT-US95-07201-43	Sequence 2, Appl
	36	252.2	8.9	22484	4	US-09-875-223-2	Sequence 1, Appl
	37	252	8.9	11531	1	US-08-068-945A-1	Sequence 1, Appl
	38	252	8.9	11531	1	US-08-442-806-1	Sequence 1, Appl
	39	249.6	8.9	2923	2	US-08-480-449-1	Sequence 1, Appl
	40	249.6	8.9	2923	1	US-08-660-542-1	Sequence 1, Appl
	41	249.6	8.9	2923	4	US-08-479-603-1	Sequence 1, Appl
	42	249.6	8.9	2927	4	US-09-232-878-5	Sequence 5, Appl
	43	249.6	8.9	81001	4	US-09-750-580-1	Sequence 1, Appl
	44	248.6	8.8	70000	4	US-09-851-896-3	Sequence 3, Appl
	45	247.6	8.8	246240	2	US-08-724-394A-20	Sequence 20, Appl

RESULT 1
US-09-817-180-3
; Sequence 3, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinlu et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001183
; CURRENT APPLICATION NUMBER: US/09/817,180
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 15297
; TYPE: DNA
; ORGANISM: Human
US-09-817-180-3

Query Match 9.9%; Score 280; DB 4; Length 15297;
Best Local Similarity 74.9%; Pred. No. 3.9e-48;
Matches 370; Conservative 0; Mismatches 110; Indels 14; Gaps 1;

QY	1641	ATTAACATATATAGAGAGTTGTGACAAACAGAAATGATAAGCTGGCAACCGTGGCACAC	1700
Db	11651	ACTCCCTTAATGCCAACCTTCCACCAGGAGAGTAAGTAACCTGGCCAGTTGCTCAC	11710
QY	1701	GCTCATAGTCTTAGCTGCTTGGAGGTTGAGGAGGAGGATGCTTGAACACAGGTGTTTC	1760
Db	11711	GCCTGTTCATCCACAGCACTTTGGGAGGCTGAGCTGGGTGATCATCTTGAGCCAGGAGTTC	11770
QY	1761	AAGCCAGCTGGGCAACATACAAAGATCCTGTCTCAAAAAAATAAATAAATAAATAAATAA	1820
Db	11771	AGATCAGCTTGGACACACAGTGAACCTCATCTGTACAAAAATACAAAAATA	11825
QY	1821	AGAGAGAGGCGGCGGTGGTGGCTCAGCCCTGTATCCAGCACTTTGGGAGGCGGAG	1880
Db	11826	-----GACTGGGCGGCTGCTCACACCTGTATCCAGCACTTTGGGAGGCGGAG	11876
QY	1881	CGGGGCGGATCAGCTGTGTCAGGAGTTTGGAGACAGCTGGCCACATCGGCAAAACCCC	1940
Db	11877	GCAGGTGGATCAGCTGTGTCAGGAGTTTGGAGACAGCTGGCCACATCGGCAAAACCCC	11936
QY	1941	GTCGTACTCAAAATGCAAAATAGCCAGCGGTGTAGCAGGCACTGTAAATCCCAAGCT	2000
Db	11937	ATCTCTACATAAAATACAAAATAGCCAGGATGGTGGCACGTGCTGTAAATCCCAAGCT	11996
QY	2001	ACTTGGGAGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTTCAGTAGCT	2060

ALIGNMENTS

```

Db      11997 ACTTGGAGGCTGAGGTGGAGAAATTCCTTGAACCCAGGAGCGGAGGCTGCAGTGAGCC 12056
Qy      2061 GAGATCGTCGCTGGTCCAGCTCCAGCTGGGCGGCAAGCAAGACTCTGTCTCAGAAAAA 2120
Db      12057 GAGATTGTGCATCTGCACTCCAGCTCCAGCTGGGCGGCAAGAGTGAACCTCATCTCAAAAAA 12116
Qy      2121 AAAAAAAGAGAGA 2134
Db      12117 CCAAAAAACAAAA 12130

RESULT 2
US-09-078-294-7/c
; Sequence 7, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 11811
; TYPE: DNA
; ORGANISM: BAC-F2 contig 3
US-09-078-294-7

Query Match          9.7%; Score 274.2; DB 4; Length 11811;
Best Local Similarity 79.7%; Pred. No. 5.5e-47;
Matches 350; Conservative 0; Mismatches 83; Indels 6; Gaps 2;

Qy      1692 GTGGCACAGCTCATAGTTCTAGCTGTGGAGTTGAGGAGGAGGATGGCTTGAACA 1751
Db      6663 GTGGCTCATCTGTATATCCAGCAGCTTTGGGAGCGGAGGTGGGAGATGACTTGAGCT 6604
Qy      1752 CAGGTGTTCAAGGCGAGCTGGGCAACATACACAGATCTCTCTCAAAAAA 1811
Db      6603 CAGAAGTTCAAGACCGCTGGCCACACACAGTGAAGCTCATCTCTACAAAATACATTTA 6544
Qy      1812 AAAAAAAGAGAGAGAGGCGGCGTGTGGTCAAGCTGTATATCCAGCAGCTTTGG 1871
Db      6543 AAGTTAG-----CTGGGCGCAGGTGTGTGGCGCAGCTGTATATCCAGCAGCTTTGG 6489
Qy      1872 GAGCGGAGCGGCGGATCACCTGTGCTCAGGAGTTGAGACAGCTGGCCCAACATGG 1931
Db      6488 GAGTCAAGTGTGGTGGATCACCTGAGTTCAGGAGTTTCGAGCCAGCTGGCCCAACATGG 6429
Qy      1932 CAAACCCGCTGTCTACTCAAAATGC-AAAAATTAGCCAGGCGTGTAGCAGGCACTGT 1990
Db      6428 TGAACCCCATCTCTACTAAATACAAAATATAGCCAGGCGTGTGGCGGCGCTGT 6369
Qy      1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGTCCCTTGAACCCAGGAGTGGAGGTT 2050
Db      6368 AATCCAGCTACTTGGAGGCTGAGGAGGAGTCCCTTGAACCCAGGAGGAGT 6309
Qy      2051 GCAGTAACTGAGTGTGCGCTTGCACCTCCAGCTGGCGGCAAGCAAGACTCTGTC 2110
Db      6308 GCAGTGAAGCAAGATCATGCCATTGCACCTCCAGCTGGCGGCAAGCAAGCAAACTCTCAA 6249
Qy      2111 TCAGAAAAA 2129
Db      6248 AAAAAA 6230

RESULT 3
US-09-735-934A-3/c
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:

```

```

; APPLICANT: Li, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO00851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

Query Match          9.6%; Score 269.8; DB 4; Length 43950;
Best Local Similarity 77.9%; Pred. No. 6.1e-46;
Matches 339; Conservative 0; Mismatches 92; Indels 4; Gaps 1;

Qy      1685 GCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGC 1744
Db      17761 GGGCACAGTGGCACACACCTGTATCCAGAACTTTGGGAGGCTGAGGTGGAGGATTC 17702
Qy      1745 TTGAACACAGGTGTTCAAGGCCAGCCCTGGGCAACATAAAGATCCTCTCTCAAAAAA 1804
Db      17701 TTGAGCGCAGAAGTTCAAGACCACCCCTGGGCAACATAGCAAGACCCCTATCTTAAAAA 17642
Qy      1805 AAAAAA 1860
Db      17641 AAAAAA 17582
Qy      1861 CAGCACTTTGGAGGCGGAGCGGCGGATACCTGTGTGTCAGGAGTTTGAGACCAAGCT 1920
Db      17581 CAGCACTTTGGAGGCGGAGCGGAGGATACCTGTGTGTCAGGAGTTTGAGACCAAGCT 17522
Qy      1921 GGCCACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATATAGCCAGGCGTGTAGC 1980
Db      17521 GACCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATATAGCCAGGCGTGTAGC 17462
Qy      1981 AGCACTCTGTAAATCCAGCTACTTTGGAGGCTGAGCGAGAGAGTGGCTGGGCGTGGC 2040
Db      17461 GGTGCTGTAGTCCAGCTACTTTGGAGGCTGAGCGAGAGAGTGGCTGGGCGTGGC 17402
Qy      2041 GGTGAGGTTGAGTAAAGTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
Db      17401 GACAGAGTTGCAATGAAGTGAATTTGATCATCTGCTGCTGCTGCTGCTGCTGCTGCT 17342
Qy      2101 AGACTCTGTCTCAGA 2115
Db      17341 AAACGCGCTCAAAA 17327

RESULT 4
US-09-798-096-10/c
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-798-096-10

Query Match          9.6%; Score 269.2; DB 4; Length 99500;
Best Local Similarity 73.7%; Pred. No. 1e-45;

```


RESULT 6

```

RESULT 6
US-09-978-197-3/c
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001178DIV
; CURRENT APPLICATION NUMBER: US/09/978,197
; CURRENT FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-978-197-3

```

RESULT 5

```

US-09-813-817-3/C
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001178
; CURRENT APPLICATION NUMBER: US/09/813,817
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-813-817-3

```

	Query Match	9.5%;	Score 267.4;	DB 4;	Length 59065;
	Best Local Similarity	78.2%;	Pred. No. 2e-45;		
	Matches 351;	Conservative 0;	Mismatches 86;	Indels 12;	Gaps 2;
QY	1680	AAAGCTGCGAACCGTGGCACACGCTCATATGCTTCTAGCTGCTTGGGAGGTTGAGGAGGGAGG	1739		
Db	23369	AGGCCAGGCATAGTGGCTCATGCTTATATCCAGC-ATTGAGAGGCCAAGGAGGGAGG	23311		
QY	1740	ATGGGCTGAACACAGAGTGTTCAAGGCCAGCCTGGGCAACATACAAAGATCCTGTCTCTCA	1799		
Db	23310	ACTGCTTGAAGCCAGAAGTTTCAGAGCACCAGCCTGGGTAAACAACCAAGACCCTATCTCTAC	23251		
QY	1800	AAAAAAAAAAAAAAAAAAGAGAGAGCGGCCGGCGTGGTGGCTCACGCCCTGTAATC	1859		

RESULT 5

Query Match	9.5%;	Score 267.4;	DB 4;	Length 59065;
Best Local Similarity	78.2%;	Pred. No. 2e-45;		
Matches 351;	Conservative 0;	Mismatches 86;	Indels 12;	Gaps 2;
Qy 1680	AAGCTGCGAACCGTGGCACACGCTCATAGTCTTAGCTGCTGGGAGGTTGAGGAGGAGG	1739		
Db	23369	AGGCCAGGCATATGGCTCATGCTATATAATCCAGC-ATTCGAGAGGCCAAGGAGGAGG	23311	
Qy 1740	ATGCGTTGACACAGCGTTTCAGGCCAGCGCTGGGCAACATACAAAGATCCTGTCTCTCA	1799		
Db	23310	ACTGCTTGAGCCAGGAAGTTTGAGACCAGCCTGGGTATCAACACCAAGACCCTATCTCTAC	23251	
Qy 1800	AAAAAAAAAAAAAAAAAAGAGAGAGAGGGCCGGCGCTGGTGGCTCACGCCTGTAATC	1859		
Db	23250	TAAAAAAATCAGTCAATCA-----GGCCAGTCATGTGGCTCACGCCTGTAATC	23202	
Qy 1860	CCAGCACTTTGGGAGGCCGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGACACAGCC	1919		
Db	23201	CCAGCACTTTGGAAGGCTCAGCGCGGCAGTCACTTGAGTTTGGAAATTCGAGATAGCC	23142	
Qy 1920	TGGCCAACTGGCAAAACCCCGTCTACTCAAAATGCAAAAATTAGCAGCGCTGGTAG	1979		
Db	23141	TGACCAACATGGAGAAACCTGTCTCTACTAAAAATACAAAATTAGCAGCGCTGGTAG	23082	
Qy 1980	CAGGCACCTGTAATCCACGCTACTTTGGGAGGCTGAGGCAGGAGAAATTCGTTGAACCCAGG	2039		
Db	23081	CACATGCTGTAATCCACGCTACTTCGAGAGGCTGAGACAGGAGAAATTCGTTGAACCCAGG	23022	

:	ORGANISM:	Homo sapiens
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(72)..(287)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(1280)..(1579)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(1819)..(1854)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(2466)..(2555)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(2863)..(2940)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(3887)..(3952)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(4896)..(4994)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(5846)..(5986)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(6965)..(7138)
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	(8556)..(8639)
:	FEATURE:	
:	NAME/KEY:	misc_feature
:	LOCATION:	(3409)
:	OTHER INFORMATION:	n = adenine or guanine or cytosine or thymine
:	FEATURE:	
:	NAME/KEY:	misc_feature
:	LOCATION:	(9214)
:	OTHER INFORMATION:	n = adenine or guanine or cytosine or thymine
:	FEATURE:	
:	NAME/KEY:	misc_feature
:	LOCATION:	(9303)
:	OTHER INFORMATION:	n = adenine or guanine or cytosine or thymine
:	FEATURE:	
:	NAME/KEY:	misc_feature
:	LOCATION:	(9311)
:	OTHER INFORMATION:	n = adenine or guanine or cytosine or thymine
:	US-09-370-265-8	
Query Match 9.4%; Score 265.6; DB 4; Length 9365;		
Best Local Similarity 79.0%; Pred. No. 2.9e-45;		
Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;		
Qy	1765	CCAGCCTGGGCACATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAGAAG- 1823
Db	7802	CCAGCATGGTGACAGAGCAAGACCCTGTCATAAAAAAAAAADGAAAAAAAG 7743
Qy	1824	-----AGAGAGGCCCGCGTGGTGCTCACGCTGTAAATCCCAGCACTTTGGGAGG 1875 : : : : :
Db	7742	AAMCGAMCAAAAAGCYRGGMGTGGTGCTCAIRCCTRTAATYCYAGCACTTTGGAGG 7683
Qy	1876	CCGAGCGGGCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGCGAAA 1935 : : : : : : :
Db	7682	CYGAGGTGGTGATCACCTGAGTCAGRAGTTRAGATTAGCCTGGCCAAACATGGTGAA 7623
Qy	1936	ACCCGCTCTGTACTAAAAATGC-AAAAATTACCCAGCGGTGTAGCAGGACCTGTGAAT 1993
Db	7622	ACCCGCTCTACTAAAAATACAAAAAATTAGCCRGCGTGTGGCGCRYGCCTGTGAAT 7563
Qy	1994	CCAGCTACTTTGGGAGGCTGAGGCAGAGAAATCGCTTGAACCCAGGAGTGGAGTTGCA 2053
Db	7562	CCACGCTACTPYGGGAGGCTGAGGCAGRAGAAATCGCTTGAACCCAGGAGGYGRAGVTGCA 7503

```
QY 2054 GTRAGCTGAGATCGTCCGCTTGCACTCCAGCCTGGCGACAAAGAGCAAGACTCTGTCTCA 2113
Db 7502 GTGAGCYGAGATCRYGCAATGCACTCCAGCCYRGMAACAAGAGCAGACTCCGTCTCA 7443
QY 2114 GAAAAAATAATAATAA 2129
Db 7442 AAAAAATAATAATAA 7427

RESULT 10
US-09-608-285A-42/c
; Sequence 42, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 14747
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13641)
; OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
US-09-608-285A-42

Query Match 9.4%; Score 265.6; DB 4; Length 14747;
Best Local Similarity 79.08; Pred. No. 3.3e-45;
Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;

QY 1765 CCAGCCTGGCAACATACAGATCTGCTCTCAAAAAAAAAAAAAAAAAAAGAG- 1823
Db 11162 CCAGCATGGGTGACAGAGCAAGACCCCTGTCTCAAAAAAAAAAAAAAAAAAG 11103
QY 1824 -----AGAGAGGCGCGGTGGTGGTCTACGCTGTATCCAGCACTTTGGGAGG 1875
Db 11102 AAMCGAMCAAAAAGCYRGGMGTGGTGGCTCAYRCCTTATCTTATCTTGGGAGG 11043
QY 1876 CCAGCGCGCGGATCACCTGTGGTCAGGAGTTTGACAGCAGCCTGGCCAAATGCAAA 1935
Db 11042 CYGAGGCGGTGGATCACCTGTGGTCAGGAGTTTGTAGAGTTCAGGAGGAGG 10983
QY 1936 ACCCGCTCTGTACTCAAAATGC-AAAAATTAGCCAGCGGTGTAGCAGGACCTGTAT 1993
Db 10982 ACCCGCTCTGTACTCAAAATGC-AAAAATTAGCCAGCGGTGTAGCAGGACCTGTAT 10923
```

```
QY 1994 CCCAGCTACTTGGGAGGCTGAGGAGAGAAATCGTTGAACCCAGGAGGTGAGGTTGCA 2053
Db 10922 CCCAGCTACTTGGGAGGCTGAGGAGAGAAATCGTTGAACCCAGGAGGTGAGGTTGCA 10863
QY 2054 CTAAGCTGAGATCGTCCGCTTGCACTCCAGCCTGGCGCACAAAGAGCAAGACTCTGTCTCA 2113
Db 10862 GTGAGCYGAGATCRYGCAATGCACTCCAGCCYRGMAACAAGAGCAGACTCCGTCTCA 10803
QY 2114 GAAAAAATAATAATAA 2129
Db 10802 AAAAAATAATAATAA 10787

RESULT 11
US-09-608-285A-59/c
; Sequence 59, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 15977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CD39-L4/L66 Gene Sequence
; NAME/KEY: CDS
; LOCATION: (245)..(461)
; NAME/KEY: CDS
; LOCATION: (1454)..(1533)
; NAME/KEY: CDS
; LOCATION: (2734)..(2877)
; NAME/KEY: CDS
; LOCATION: (4364)..(4439)
; NAME/KEY: CDS
; LOCATION: (4679)..(4714)
; NAME/KEY: CDS
; LOCATION: (5326)..(5414)
; NAME/KEY: CDS
; LOCATION: (5723)..(5802)
; NAME/KEY: CDS
; LOCATION: (6751)..(6812)
; NAME/KEY: CDS
; LOCATION: (7758)..(7859)
```

[illegible]

US-08-688-145-1

Query Match 9.2%; Score 259.2; DB 1; Length 3885;
Best Local Similarity 73.4%; Pred. No. 4.7e-44;
Matches 372; Conservative 0; Mismatches 129; Indels 6.

1566	Qy	1566	AAACACAGATGATAAAGCTCGCAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAG	1725
1599	Db	1599	AAAAATTGAAGAAAGGGCCAGGTGTGATGGCTCATTCCTATAGTCCAGCACATTTGGGAG	1540
		1599	AAAAATTGAAGAAAGGGCCAGGTGTGATGGCTCATTCCTATAGTCCAGCACATTTGGGAG	
1726	Qy	1726	GTTTAGGAGGAGGATGTGGCTTTGAACACAGGTGTTCAAGGCCAGCTTGGGCAACATTAACAA	1785
1539	Db	1539	GCCAAAGGCAGAGGATCATCTGAAGCCAGGAGTTCAAGGACGACCTGGG---CTGTACCA	1483
		1539	GCCAAAGGCAGAGGATCATCTGAAGCCAGGAGTTCAAGGACGACCTGGG---CTGTACCA	
1786	Qy	1786	GATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAGAGAGAGAGGCGCGGCGCTGTGGC	1845
1482	Db	1482	TGGTGAAGACCTGTGTCAACTAAAAAACAAAGAGGAGGGGGCGCGCTGTGTAGC	1423
		1482	TGGTGAAGACCTGTGTCAACTAAAAAACAAAGAGGAGGGGGCGCGCTGTGTAGC	
1846	Qy	1846	TCACGCGTGTAAATCCAGCACATTTGGAGGCGCCAGCGCGCGGATCACCTGTGGTCAGGA	1905
1422	Db	1422	TCACACTGTAAATACCAACGTGTGGAGGCGCGAAGTGGGTGTATCATCTGTAGGTCAGGA	1363
		1422	TCACACTGTAAATACCAACGTGTGGAGGCGCGAAGTGGGTGTATCATCTGTAGGTCAGGA	
1906	Qy	1906	GTTTGAACACAGCCTGGCCAACTGGCAAAACCCCGTCTGTACTCAAAATGC--AAAAAT	1963
1362	Db	1362	GTTTGTGACGAGCCTAGCCAACTAGCGAAACCCCGTCTCCACTAAAAATACAAAAAAT	1303
		1362	GTTTGTGACGAGCCTAGCCAACTAGCGAAACCCCGTCTCCACTAAAAATACAAAAAAT	
1964	Qy	1964	TAGCCAGGCGTGTGTAGCAGCACCTGTAAATCCAGCTACTTGGAGGCGTGTAGCGCAGGA	2023
1302	Db	1302	TAGCTGGAGTGTGGCGGTGCTGTGAGTCCAGCTGCTTGGAGGCGTGTAGCGCAGGA	1243
		1302	TAGCTGGAGTGTGGCGGTGCTGTGAGTCCAGCTGCTTGGAGGCGTGTAGCGCAGGA	
2024	Qy	2024	ATCGCTTGAAACCCAGGAGTGGAGTTGTCAGTAAAGCTGAGATCGTCCCGCTTGACATCCAG	2083
1242	Db	1242	ATCGCTTGAAACCCAGGAGGAGGTTGTCAGTGTAGCCTAGGTCAGTGTAGCCTAGG	1183
		1242	ATCGCTTGAAACCCAGGAGGAGGTTGTCAGTGTAGCCTAGGTCAGTGTAGCCTAGG	
2084	Qy	2084	CTGTGGCGCAAGAGCAAGACTCTGTCTCAGAAAAAANAANAAGAGAGAGAGAGAG	2143
1182	Db	1182	CTGTGGCAAC-AGAGCAAGCCTCTATCTNNAAAAAANAANAAGAAAGGAGCAGAAA	1124
		1182	CTGTGGCAAC-AGAGCAAGCCTCTATCTNNAAAAAANAANAAGAAAGGAGCAGAAA	
2144	Qy	2144	AAAGAGAACAAATATTTGGGAGAGAAGG	2170
1123	Db	1123	AAAAAGAAATTAAGCTGGATGTGGTGG	1097
		1123	AAAAAGAAATTAAGCTGGATGTGGTGG	

RESULT 14
US-09-735-934A-3
; Sequence 3, Application US/09735934A

```

; Patent No. 6372468
;
; GENERAL INFORMATION:
;
; APPLICANT: LI, JIAYIN et al
;
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;
; TITLE OF INVENTION: THEREOF
;
; FILE REFERENCE: CL000851
;
; CURRENT APPLICATION NUMBER: US/09/735,934A
;
; CURRENT FILING DATE: 2000-12-14
;
; NUMBER OF SEQ ID NOS: 4
;
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3

```

```

: LENGTH: 43950
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-735-934A-3

```

	Query Match	9.2%	Score 258.6;	DB 4;	Length 43950;
	Best Local Similarity	78.8%;	Pred. No. 1.2e-43;		
	Matches 346;	Conservative 0;	Mismatches 89;	Indels 4;	Gaps 3;
QY	1696	CACACGCTCATAGTCTTAGCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTCAACACAGG	1755		
Db	35148	CTCAGCGCTGTAATCCACGATTTTGGGAGGCCAAGGTGGGTGGATCTACTTGGTCCAGG	35207		
QY	1756	TGTTTCAGGCCGACGCTGGCGCATACATCAAGAATCCCTGTCTCTCAAAAAA	1815		

Db	35208	AGTTTGAGACCGCCTGGTCAACATGGTGAACCCCTGTCTCTACTATAAAATAAAAA	35267
Qy	1816	AAAGAAAGAGAGAGGGCGGGCGCTGGTGGCTACGCGCTGTAATCCAGCACCTTGGGAGG	1875
Db	35268	AAATTATCAGACGACGAGTCTGGGCCGGTGGCTACGCGCTGTAATCCAGCACCTTGGGAGG	35327
Qy	1876	CCGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGACGACGCTGGCCAACTGGCAAA	1935
Db	35328	CCGAGGTGGGGAATCAC--GAGTCAAGAGTTTGAGACCGCTGGCCAACTGGTGA	35385
Qy	1936	ACCCCGTCTCTACTCAAAAT-GCAAAATTTAGCAGGCGTGGTAGACGACCTCTGAATC	1994
Db	35386	ACCCCATCTCTACTAAAAATAGAAAAATTTAGCTGGCATAGTGGCCAGCGCTCTGAATC	35445
Qy	1995	CCAGCTACTTTGGGAGCCTGAGGCAGAGAGAACTCGCTTGAACCCAGAGAGGTGGAGTTGCAG	2054
Db	35446	CCAGCTACTCGGAGGCTGAGGCAGAGAGACTCACTTGAAGCCCTGGAGGTAGAGTTGCAG	35505
Qy	2055	TAAGCTGAGATCGTGGCTTGGCACTCCAGCGCTGGCGACAGCAAGCTCTGCTCTCAG	2114
Db	35506	TGAGCCGAGATCGTGCCACTGCACTCCAGCCAGGGAC-AGAGTGACACTCCGCGCTCA	35564
Qy	2115	AAAAAAAAAAAAAAAAAGAG	2133
Db	35565	AAAAAAAAAAAAAAAAATTAG	35583

RESULT 15

```

US-09-798-096-10
: Sequence 10, Application US/09798096
: Patent No. 6399378
: GENERAL INFORMATION:
: APPLICANT: Donna T. Ward
: APPLICANT: Andrew T. Watt
: TITLE OF INVENTION: ANTISENSE MODULATION
: FILE REFERENCE: RTS-0207
: CURRENT APPLICATION NUMBER: US/09/798,096
: CURRENT FILING DATE: 2001-03-01
: NUMBER OF SEQ ID NOS: 89
: SEQ ID NO 10
: LENGTH: 99500
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
US-09-798-096-10

```

Query Match	9.1%	Score 257.6;	DB 4;	Length 99500;
Best Local Similarity	69.1%;	Pred. No. 2.3e-43;		
Matches 376;	Conservative 0;	Mismatches 149;	Indels 19;	Gaps 1;
Qy	1616	TGTGTGATGCTTCTCCCAAGGTATATTAACATATATAAGAGAGTTGTGTGACAAACAGAAT	1675	
Db	12586	TGCTTAGTTTTCTATAGTGCTGTAATAAATACCATAAAGTCTAGTGCCTTAACAACAC	12645	
Qy	1676	GATAAAGTCGCAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGGTGAGGAGG	1735	
Db	12646	ATAGAGCGCCGGGAACATTGGCTCATCCCTGTATCTACGAGTTTGGGAGGCCAAGGTGG	12705	
Qy	1736	GAGGATGGCTTGAACACAGGTGTTCAGGCCACGCTGGGCAACATAAACAGATCCTTGCT	1795	
Db	12706	GAGTATCGCGGGAGCCAGAAAATTCGAGACCAGTCTGGGCAACACGGTGAGTCCCTGTCT	12765	

Qy	1796	CTCAAAAAAAAAAAAAAAAAAAGAGAGAGCGGCGTGGTGGCTCAGCCCTGT	1855
Db	12766	CTGTTTTAAAAATAACAATAACA-----CACGGTAGCTCACACCTGT	12806
Qy	1856	AATCCAGACATTTGGGAGCGGAGCGCGCGGATCACCTGTGTCAGGAGTTTGGAGACC	1915
Db	12807	AATCCAGACATTTGAGAGCGGAGCGCGGAGTTCAGCTGAGGTTCGGGAGCTTTGAGACT	12866
Qy	1916	AGCCTGGCCACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATATAGCCAGGGGTG	1975
Db	12867	AGCCTGGCCACGTGGGAAACTTCACTCTACTAAAAATACAAAATATAGCCAGGGATG	12926

Search completed: June 19, 2003, 11:13:50
Job time : 122.35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:06:39 ; Search time 378.09 Seconds
(without alignments)
10937.110 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 2818
Sequence: 1 agagaccatcattatga.....aaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2818	100.0	2818	9	US-09-966-880A-7
2	2174.6	77.2	6564	9	Sequence 7, Appli
3	2174.6	77.2	11204	9	Sequence 10, Appl
4	2172	77.1	2172	9	Sequence 35, Appl
5	603.4	21.4	2440	9	Sequence 15, Appl
6	295.2	10.5	13862	9	Sequence 1, Appli
7	295.2	10.5	13862	9	Sequence 5477, Ap
8	295.2	10.5	13862	9	Sequence 10204, A
9	292.6	10.4	7809	9	Sequence 1003, Ap
10	284.4	10.1	32249	9	Sequence 6094, Ap
11	281.4	10.0	65608	9	Sequence 7364, Ap
12	281.4	10.0	65608	10	Sequence 180, App
13	281.4	10.0	65608	10	Sequence 292, App
14	280.2	9.9	11821	10	Sequence 119, App
15	280	9.9	15297	9	Sequence 2857, Ap
16	278.2	9.9	32204	9	Sequence 3, Appli
17	275.4	9.8	3966	9	Sequence 8537, Ap
18	275.4	9.8	3966	9	Sequence 227, App
19	274.8	9.8	15362	10	Sequence 227, App
					Sequence 2856, Ap

c	20	274.2	9.7	11811	9	US-09-728-552-7	Sequence 7, Appli
	21	274	9.7	4963	10	US-09-764-877-2903	Sequence 2903, Ap
	22	273.6	9.7	174424	10	US-09-967-768A-314	Sequence 314, App
	23	273	9.7	26657	10	US-09-810-673A-3	Sequence 3, Appli
	24	271	9.6	271	9	US-09-966-880A-13	Sequence 13, Appli
c	25	270.2	9.6	17286	10	US-09-764-877-3234	Sequence 3234, Ap
	26	269.8	9.6	43950	12	US-10-060-332-3	Sequence 3, Appli
	27	269.4	9.6	16086	10	US-09-764-877-2385	Sequence 2385, Ap
c	28	268	9.5	11677	9	US-10-082-830-134	Sequence 134, App
	29	267.8	9.5	35641	10	US-09-962-436-306	Sequence 306, App
	30	267.8	9.5	35641	10	US-09-880-107-2225	Sequence 2225, Ap
c	31	266	9.4	56737	10	US-09-782-378A-17	Sequence 17, Appli
	32	266	9.4	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c	33	265.6	9.4	9365	9	US-10-092-063-8	Sequence 8, Appli
	34	265.6	9.4	9365	12	US-10-091-085-8	Sequence 8, Appli
	35	265	9.4	9903	9	US-09-764-891-8188	Sequence 8188, Ap
	36	264.8	9.4	32184	9	US-09-764-891-8538	Sequence 8538, Ap
	37	264.8	9.4	57130	10	US-09-835-081-3	Sequence 79, Appli
c	38	264.4	9.4	145831	10	US-09-969-708-79	Sequence 2116, Ap
	39	264.4	9.4	145831	10	US-09-954-456-2116	Sequence 2116, Ap
c	40	264	9.4	12452	9	US-09-764-877-3989	Sequence 3989, Ap
	41	263.6	9.4	4767	9	US-09-764-891-5474	Sequence 5474, Ap
c	42	263	9.3	10838	10	US-09-764-877-2698	Sequence 2698, Ap
	43	263	9.3	13608	10	US-09-861-846-3	Sequence 3, Appli
c	44	263	9.3	38374	10	US-09-880-107-3463	Sequence 3463, Ap
	45	262.8	9.3	3740	9	US-09-764-891-9984	Sequence 9984, Ap

ALIGNMENTS

RESULT 1
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Query Match 100.0%; Score 2818; DB 9; Length 2818;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2818; Conservative 0; Mismatches 0; Gaps 0;
QY 1 AGAGAACCATCATTATTAAGTGAAGTCTGGCCCTGAGACTTTCGAGGAGGCAAGA 60
|||||

Db 1 AGAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGA 60
QY 61 AGACACTCTGACACCACTATPGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCA 120
Db 61 AGACACTCTGACACCACTATPGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCA 120
QY 121 ATTCAAAAATCTCCGCTGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAA 180
Db 121 ATTCAAAAATCTCCGCTGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAA 180
QY 181 GAGCGGTGACAGTGCTACATCCCTTTTTCACATGGACTTTGGTTATCTTCGCAATAAGAAGG 240
Db 181 GAGCGGTGACAGTGCTACATCCCTTTTTCACATGGACTTTGGTTATCTTCGCAATAAGAAGG 240
QY 241 CTGCCAGTGGAAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCG 300
Db 241 CTGCCAGTGGAAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCG 300
QY 301 CTGCTACCGGTCACCTGGTTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGT 360
Db 301 CTGCTACCGGTCACCTGGTTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGT 360
QY 361 GGCGGACTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCNCTA 420
Db 361 GGCGGACTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCCCTTA 420
QY 421 CTCTGTGAGACCGCAAGGCTGAGCCGAGGGCTCGCGGCTGACCGCGCGGGGT 480
Db 421 CTCTGTGAGACCGCAAGGCTGAGCCGAGGGCTCGCGGCTGACCGCGCGGGGT 480
QY 481 GCAAAATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATCTTTGTAGAAAA 540
Db 481 GCAAAATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATCTTTGTAGAAAA 540
QY 541 CCATGAAAGAACTTTCAAGCCCTGGGAAGGCTGCATGAAAAATTCAGTTCTCTCCAG 600
Db 541 CCATGAAAGAACTTTCAAGCCCTGGGAAGGCTGCATGAAAAATTCAGTTCTCTCCAG 600
QY 601 ACAGCTTGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCG 660
Db 601 ACAGCTTGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCG 660
QY 661 TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTG 720
Db 661 TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTG 720
QY 721 AAGACAGTGGATAAAAAACAGTCCCTCAAGTCTTCTCTGTTTATTTTCAACTCTCAC 780
Db 721 AAGACAGTGGATAAAAAACAGTCCCTCAAGTCTTCTCTGTTTATTTTCAACTCTCAC 780
QY 781 TTTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTG 840
Db 781 TTTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTG 840
QY 841 AAAATAGAGAAGAACACAGGCTCGCCAGGGACGTGCTGCAATTTGGTCAGTTTGAAT 900
Db 841 AAAATAGAGAAGAACACAGGCTCGCCAGGGACGTGCTGCAATTTGGTCAGTTTGAAT 900
QY 901 GCAACATTTGCCCTACTGGGATAACAGAACTGCAAGACCTGGGAGCATCTTAAAGTGT 960
Db 901 GCAACATTTGCCCTACTGGGATAACAGAACTGCAAGACCTGGGAGCATCTTAAAGTGT 960
QY 961 CAACGTTTTCTATGACTTTTAGTAGGATGAGAGCAGAGGTAGATCCATAAAGCATG 1020
Db 961 CAACGTTTTCTATGACTTTTAGTAGGATGAGAGCAGAGGTAGATCCATAAAGCATG 1020
QY 1021 GTGAGAGGATCAAACTGTTTTTATCAACATCTTTATTTATTTGATTCATTCAGTTAAC 1080
Db 1021 GTGAGAGGATCAAACTGTTTTTATCAACATCTTTATTTATTTGATTCATTCAGTTAAC 1080
QY 1081 AGTGGTGTAGTGATAGATTTTCTATTTCTTTTCCCTTGACGCTTTACTTTCAAGTAACAC 1140
Db 1081 AGTGGTGTAGTGATAGATTTTCTATTTCTTTTCCCTTGACGCTTTACTTTCAAGTAACAC 1140

QY 1141 AAACCTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTTG 1200
Db 1141 AAACCTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTTG 1200
QY 1201 GACCCCAACCATCTCTCCAAAGCATTAATATCAATCATCGCTGTATGTTTTAAATCAG 1260
Db 1201 GACCCCAACCATCTCTCCAAAGCATTAATATCAATCATCGCTGTATGTTTTAAATCAG 1260
QY 1261 CAGAGCATGTTTTTATGTTGTACAAAGAAAGATTGTTATGGTGGGGATGGAGGTATA 1320
Db 1261 CAGAGCATGTTTTTATGTTGTACAAAGAAAGATTGTTATGGTGGGGATGGAGGTATA 1320
QY 1321 GACCATGCATGGTCACCTTCAAGCTACTTTTAAATAAGGATCTTAAATGGGCAGGAGAC 1380
Db 1321 GACCATGCATGGTCACCTTCAAGCTACTTTTAAATAAGGATCTTAAATGGGCAGGAGAC 1380
QY 1381 TGTGAACAAGACACCCCTAAATAATGGTGTGATCTGAAGTAGCAAACTTCTGGAACGC 1440
Db 1381 TGTGAACAAGACACCCCTAAATAATGGTGTGATCTGAAGTAGCAAACTTCTGGAACGC 1440
QY 1441 AAACCTCTTTAAGGAAGTCCCTTAATTTAGAAACACCCCAAACTTCACATATCATATA 1500
Db 1441 AAACCTCTTTAAGGAAGTCCCTTAATTTAGAAACACCCCAAACTTCACATATCATATA 1500
QY 1501 GCAAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAAAATCTATTGGCTCTCGT 1560
Db 1501 GCAAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAAAATCTATTGGCTCTCGT 1560
QY 1561 GGGTCTCTCATCTCAGAAAATGCCAATCAGTCAAGGTTTGTACATTTGTATGTGTGT 1620
Db 1561 GGGTCTCTCATCTCAGAAAATGCCAATCAGTCAAGGTTTGTACATTTGTATGTGTGT 1620
QY 1621 GATGCTTCTCCCAAGGATATTAACCTATATAGAGAGTGTGTGACAAACAGAAATGATA 1680
Db 1621 GATGCTTCTCCCAAGGATATTAACCTATATAGAGAGTGTGTGACAAACAGAAATGATA 1680
QY 1681 AGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTCTTGGGAGGTTGAGGAGGGAGGA 1740
Db 1681 AGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTCTTGGGAGGTTGAGGAGGGAGGA 1740
QY 1741 TGGCTTGAACACAGGTTTCAAGGCCACGCTGGGCAACATAACAAAGATCTGTCTCTCAA 1800
Db 1741 TGGCTTGAACACAGGTTTCAAGGCCACGCTGGGCAACATAACAAAGATCTGTCTCTCAA 1800
QY 1801 AAAAAAATAAAAAAAGAGAGAGGCGCGGTGCTGCTCACGCTCTAATCC 1860
Db 1801 AAAAAAATAAAAAAAGAGAGAGGCGCGGTGCTGCTCACGCTCTAATCC 1860
QY 1861 CAGCAGCTTTGGGAGGCGGAGCGGCGGATCAGCTGTGGTCAAGAGTTTGAGACGAGCCT 1920
Db 1861 CAGCAGCTTTGGGAGGCGGAGCGGCGGATCAGCTGTGGTCAAGAGTTTGAGACGAGCCT 1920
QY 1921 GGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGCGGTGAGC 1980
Db 1921 GGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGCGGTGAGC 1980
QY 1981 AGGCAGCTGTAAATCCAGCTACTTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGA 2040
Db 1981 AGGCAGCTGTAAATCCAGCTACTTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGA 2040
QY 2041 GGTGGAGGTTGCAAGTAAAGTGTAGATCGTGGCGTTGCACTCCAGCCTGGGCGACAAGAGCA 2100
Db 2041 GGTGGAGGTTGCAAGTAAAGTGTAGATCGTGGCGTTGCACTCCAGCCTGGGCGACAAGAGCA 2100
QY 2101 AGACTCTGTCTCAGAAAAAATAAAAAAAGAGAGAGAGAGAAACAAATATTTG 2160
Db 2101 AGACTCTGTCTCAGAAAAAATAAAAAAAGAGAGAGAGAGAAACAAATATTTG 2160
QY 2161 GGAGAGAAGGATGGGGAAGCATTCGAAGGAAATTTGCTTTATCCCAACAAAATGTAAAGGA 2220
Db 2161 GGAGAGAAGGATGGGGAAGCATTCGAAGGAAATTTGCTTTATCCCAACAAAATGTAAAGGA 2220

QY 2221 GCCAATAAGGGATCCCTATTTGTCCTCTTTGGTGCTATTTGTCCTCCTAACAACTGCTCTT 2280
Db |||||
QY 2221 GCCAATAAGGGATCCCTATTTGTCCTCTTTGGTGCTATTTGTCCTCCTAACAACTGCTCTT 2280
Db |||||
QY 2281 GACAGTGAGAAAATATTCAGAAATACCATATCCCTGCGCGTTATTTACCTAGCAACCGT 2340
Db |||||
QY 2281 GACAGTGAGAAAATATTCAGAAATACCATATCCCTGCGCGTTATTTACCTAGCAACCGT 2340
Db |||||
QY 2341 TGCAATGAAGATGAGCAGATCCACAGGAAAACCTTGAATGCACAACTGCTCTTATTTTAATC 2400
Db |||||
QY 2341 TGCAATGAAGATGAGCAGATCCACAGGAAAACCTTGAATGCACAACTGCTCTTATTTTAATC 2400
Db |||||
QY 2401 TTATTTGACATAAGTTTGTAAAGAGTTAAATAATTTGTTACTTCATGTAATTTATATTT 2460
Db |||||
QY 2401 TTATTTGACATAAGTTTGTAAAGAGTTAAATAATTTGTTACTTCATGTAATTTATATTT 2460
Db |||||
QY 2461 TTATATTTTGGCGCTCAATGATTTTATTAACATGATTTCCCTTTCTGATATATTTGA 2520
Db |||||
QY 2461 TTATATTTTGGCGCTCAATGATTTTATTAACATGATTTCCCTTTCTGATATATTTGA 2520
Db |||||
QY 2521 AATGAGTCTCAAAAGCTTCATAAAATTTATAACTTTAGAAATGATTTCTTAATAACAACGTAT 2580
Db |||||
QY 2521 AATGAGTCTCAAAAGCTTCATAAAATTTATAACTTTAGAAATGATTTCTTAATAACAACGTAT 2580
Db |||||
QY 2581 GTAATGTAACATTCAGTAATGCTGCTACGAAGCCATTTCTCTGATTTTGTAGTAACT 2640
Db |||||
QY 2581 GTAATGTAACATTCAGTAATGCTGCTACGAAGCCATTTCTCTGATTTTGTAGTAACT 2640
Db |||||
QY 2641 TTATGACAGCAAAATTTGCTTCTGCTCCTCACTTTCAATCAGTTTAAATAGATAATAAT 2700
Db |||||
QY 2641 TTATGACAGCAAAATTTGCTTCTGCTCCTCACTTTCAATCAGTTTAAATAGATAATAAT 2700
Db |||||
QY 2701 TTTGGAAGCTGTGAAGATAAAATACCAATAAAATAATAATAAAAGTGATTTATATGAAGT 2760
Db |||||
QY 2701 TTTGGAAGCTGTGAAGATAAAATACCAATAAAATAATAATAAAAGTGATTTATATGAAGT 2760
Db |||||
QY 2761 TAAATAAAAAATCATGATGATGGAATAAACTTTGAAAAAATAAAATGAAAAA 2818
Db |||||
QY 2761 TAAATAAAAAATCATGATGATGGAATAAACTTTGAAAAAATAAAATGAAAAA 2818
Db |||||

RESULT 2

US-09-966-880A-10
; Sequence 10, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 6564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-10

Query Match 77.2%; Score 2174.6; DB 9; Length 6564;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 617 CTTTGGCCCTGTATGAGTTTGATGACTTACGAGCGCATTTTCGTAATTTGGGACTTTTGA 676

Db |||||
3735 CTACAGCCCTGTATGAGTTTGATGACTTACGAGACGCATTTTCGTACTTTGGACATTTGA 3794
QY |||||
677 TAGCAACTTCCAGGAATGTCACACAGCATGAATAATCTCTGCTGGAAGACAGTGGATAAAA 736
Db |||||
3795 TAGCAACTTCCAGGAATGTCACACAGCATGAATAATCTCTGCTGGAAGACAGTGGATAAAA 3854
QY |||||
737 AACAGTCTTCAAGTCTTCTCTGTTTTTATTTCAACTCTCTCACTTCTTCTTAGAGTTTACA 796
Db |||||
3855 AACAGTCTTCAAGTCTTCTCTGTTTTTATTTCAACTCTCTCACTTCTTCTTAGAGTTTACA 3914
QY |||||
797 GAAAAAATATTTATATACGACTCTTTTAAAGATCTATGCTTTGAAAAATAGAGAAGAAC 856
Db |||||
3915 GAAAAAATATTTATATACGACTCTTTTAAAGATCTATGCTTTGAAAAATAGAGAAGAAC 3974
QY |||||
857 ACAGGTCTGCGCAGGAGCGTCTGCAATTTGGTGCAGTTTGTGAATGCAACATTTCCCCCTA 916
Db |||||
3975 ACAGGTCTGCGCAGGAGCGTCTGCAATTTGGTGCAGTTTGTGAATGCAACATTTCCCCCTA 4034
QY |||||
917 CTGGGAATAACAGAACTGCAGGACCTGGGAGGATCCTAAAGTGTCAACGTTTTTCTATGA 976
Db |||||
4035 CTGGGAATAACAGAACTGCAGGACCTGGGAGGATCCTAAAGTGTCAACGTTTTTCTATGA 4094
QY |||||
977 CTTTGTAGTGTAGATGAGCAGAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAATG 1036
Db |||||
4095 CTTTGTAGTGTAGATGAGCAGAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAATG 4154
QY |||||
1037 TTTTATATATCAACATCCTTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1096
Db |||||
4155 TTTTATATATCAACATCCTTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 4214
QY |||||
1097 GATTTTCTTATTTTCCCTTGACGTTTCAAGTAAACAACTCTTCCATCAGG 1156
Db |||||
4215 GATTTTCTTATTTTCCCTTGACGTTTCAAGTAAACAACTCTTCCATCAGG 4274
QY |||||
1157 CCATGATCTATAGGACCTCCTAATGAGATATCTGGGTGATTTGATGATGATGATGATGATG 1216
Db |||||
4275 CCATGATCTATAGGACCTCCTAATGAGATATCTGGGTGATTTGATGATGATGATGATGATG 4334
QY |||||
1217 TCCAAAGCATTAATATCCCAATCATGCGCTGTATGTTTTTATCAGCAGAGACATGTTTTTA 1276
Db |||||
4335 TCCAAAGCATTAATATCCCAATCATGCGCTGTATGTTTTTATCAGCAGAGACATGTTTTTA 4394
QY |||||
1277 TGTGTTGTACAAAAGAAAGATTGTTATGGTGGGATGGAGTATAGACCATGATGCTAC 1336
Db |||||
4395 TGTGTTGTACAAAAGAAAGATTGTTATGGTGGGATGGAGTATAGACCATGATGCTAC 4454
QY |||||
1337 CTTCAAGCTACTTTAATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACCC 4514
Db |||||
1397 TAATAATGGGTTGATGCTGGAAGTAGCAAAATCTTCTGGAAACCCAACTCTTTTAAGGAA 1456
QY |||||
4515 TAATAATGGGTTGATGCTGGAAGTAGCAAAATCTTCTGGAAACCCAACTCTTTTAAGGAA 4574
Db |||||
1457 GTCCCTAAATTTAGAAAACCCACAACTTACATATATCATATATAGCAAAATTTGAAGG 1516
QY |||||
4575 GTCCCTAAATTTAGAAAACCCACAACTTACATATATCATATATAGCAAAATTTGAAGG 4634
Db |||||
1517 AAGTTGCTTGAATTTGGGAGAGGAAAATCTATTTGGCTCTCTGTTGGGTCTCTTCTCATCTCA 1576
QY |||||
4635 AAGTTGCTTGAATTTGGGAGAGGAAAATCTATTTGGCTCTCTGTTGGGTCTCTTCTCATCTCA 4694
Db |||||
1577 GAAATGCCCAATCAGGTCAAGGTTTGTACATTTTGTATGTTGTTGTTGTTGTTGTTGTTG 1636
QY |||||
4695 GAAATGCCCAATCAGGTCAAGGTTTGTACATTTTGTATGTTGTTGTTGTTGTTGTTGTTG 4754
Db |||||
1637 GTATATTAATATATAAGAGAGTTGTGACAAAACAGATGATATAAGCTGCGAACCGTGGC 1696
QY |||||
4755 GTATATTAATATATAAGAGAGTTGTGACAAAACAGATGATATAAGCTGCGAACCGTGGC 4814
Db |||||
1697 ACACGCTCATAGTTCTAGCTTGGGAGGTTTGGGAGGAGGATGCTTTGAACACAGGT 1756
QY |||||

1217	QY	TCCAAGACATTAATATCAAAATCATCGCGTGTATGTTTTAATACAGAGAAGCATGTTTTTA	1276
9550	Db	TCCAAGACATTAATATCAAAATCATCGCGTGTATGTTTTAATACAGAGAAGCATGTTTTTA	9609
1277	QY	TCGTTGTACAAAGAAGATGCTTATGCGTGGGGATGGAGGTATAGACCATCCATGGTCCAC	1336
9610	Db	TCGTTGTACAAAGAAGATGCTTATGCGTGGGGATGGAGGTATAGACCATCCATGGTCCAC	9669
1337	QY	CTTCAAGCTACTTTAATAAAGGATCTTAAATGGCAGGAGGACTGTGAACAACACACACC	1396
9670	Db	CTTCAAGCTACTTTAATAAAGGATCTTAAATGGCAGGAGGACTGTGAACAACACACACC	9729
1397	QY	TAATAATGGGTTGATGTCGAAGTAGCAAAATCTTCTGAAACGCAAACTCTTTTAAGGAA	1456
9730	Db	TAATAATGGGTTGATGTCGAAGTAGCAAAATCTTCTGAAACGCAAACTCTTTTAAGGAA	9789
1457	QY	GTCCCTCAATTTAGAAACACCCACAACCTTCACATATCATATAATTAGCAAACAATTTGGAAG	1516
9790	Db	GTCCCTCAATTTAGAAACACCCACAACCTTCACATATCATATAATTAGCAAACAATTTGGAAG	9849
1517	QY	AAAGTTGCTTGAATGTTGGGAGAGAAATCTATTGGCTCTCTGGGTCTCTTCATCTCA	1576
9850	Db	AAAGTTGCTTGAATGTTGGGAGAGAAATCTATTGGCTCTCTGGGTCTCTTCATCTCA	9909
1577	QY	GAATGCCAATCAGGTCGAAGTTTGCTACATTTTGTATGTGTGATGCTTCTCCCAAAG	1636
9910	Db	GAATGCCAATCAGGTCGAAGTTTGCTACATTTTGTATGTGTGATGCTTCTCCCAAAG	9969
1637	QY	GTATATTAACTATATAAGAGAGTTGTGACAAAACAGAATGATAAGCTGCGAACCGTGGC	1696
9970	Db	GTATATTAACTATATAAGAGAGTTGTGACAAAACAGAATGATAAGCTGCGAACCGTGGC	10029
1697	QY	ACACGCTCATAGTTCTAGCTGCTTGGGAGGTTTGAGGAGGAGGATGGCTTCAACACAGGT	1756
10030	Db	ACACGCTCATAGTTCTAGCTGCTTGGGAGGTTTGAGGAGGAGGATGGCTTCAACACAGGT	10089
1757	QY	GTTTCAAGGCCAGCCTTGGGCAACATAACAAGATCCTGTCTCAAAAAAATAAAAAA	1816
10090	Db	GTTTCAAGGCCAGCCTTGGGCAACATAACAAGATCCTGTCTCAAAAAAATAAAAAA	10149
1817	QY	AGAAGAAGAGAGGCCGGCGGTGTGGCTCACGCGTGTATCCAGGACATTTGGGAGGC	1876
10150	Db	AGAAGAAGAGAGGCCGGCGGTGTGGCTCACGCGTGTATCCAGGACATTTGGGAGGC	10209
1877	QY	CGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGCCAGCCTGGCCCAACATGGCAAAA	1936
10210	Db	CGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGCCAGCCTGGCCCAACATGGCAAAA	10269
1937	QY	CCCCGTGCTGTACTCAAAATGCAAAATTTAGCCAGCGGTGTAGCAGGACCTGTATATCCC	1996
10270	Db	CCCCGTGCTGTACTCAAAATGCAAAATTTAGCCAGCGGTGTAGCAGGACCTGTATATCCC	10329
1997	QY	AGCTACTTGGGAGGCTGAGGCAGGAGATTCGTTGAACCCAGGAGGTGGAGTTGCAAGTA	2056
10330	Db	AGCTACTTGGGAGGCTGAGGCAGGAGATTCGTTGAACCCAGGAGGTGGAGTTGCAAGTA	10389
2057	QY	AGCTCAGATCTGTGCGCTTGGCACTCCAGCCTGGGGACAAGAGCAAGACTCTGTCTCAGAA	2116
10390	Db	AGCTCAGATCTGTGCGCTTGGCACTCCAGCCTGGGGACAAGAGCAAGACTCTGTCTCAGAA	10449
2117	QY	AAAAAATAAAG	2176
10450	Db	AAAAAATAAAG	10509
2177	QY	AAGCATTCGAAGAAATTTGCTTTATCCACAATAATGTAAGGAGCCCAATTAAGGGATCCC	2236
10510	Db	AAGCATTCGAAGAAATTTGCTTTATCCACAATAATGTAAGGAGCCCAATTAAGGGATCCC	10569
2237	QY	TATTTGCTCTCTTTTGGTGTCTATTGTTGCCCTAAACAACCTGCTTTTGACAGTCAGAAAAA	2296
10570	Db	TATTTGCTCTCTTTTGGTGTCTATTGTTGCCCTAAACAACCTGCTTTTGACAGTCAGAAAAA	10629
2297	QY	TTTCAAGTAACCATATTCCTCTGCGCTTTATTACCTAGCAACCCCTTGAATGAAGATGAGC	2356

Db	10630	 TTCAGAAATACCATATCCCTGCGCGTATTATACCTAGCAACCCCTTGCATTAAGATGAGC	106899
Qy	2357	AGATCCACAGGAAACCTTCAATGCGCAACCTGCTTATTTTAATCTTATTGTACATAAGTT	2416
Db	10690	AGATCCACAGGAAACCTTCAATGCGCAACCTGCTTATTTTAATCTTATTGTACATAAGTT	10749
Qy	2417	TGTAAGAGAGTTTAAAAATTTGTTACTTCATGTATTCATTATATATTTATTTTGGCT	2476
Db	10750	TGTAAGAGAGTTTAAAAATTTGTTACTTCATGTATTCATTATATATTTATTTTGGCT	10809
Qy	2477	CTAATGATTTTTTATTAAACATGATTTCCCTTTCTGATATATTGAAATGGAGTCTCAAGC	2536
Db	10810	CTAATGATTTTTTATTAAACATGATTTCCCTTTCTGATATATTGAAATGGAGTCTCAAGC	10859
Qy	2537	TTTCATAAAATTTAACTTTAGAAATGATTCCTAATAACAACGCTATCTGAATTTGTAACATTC	2596
Db	10870	TTTCATAAAATTTAACTTTAGAAATGATTCCTAATAACAACGCTATCTGAATTTGTAACATTC	10929
Qy	2597	AGTAATGGTGCTACGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTTATGACAGCAAAAT	2656
Db	10930	AGTAATGGTGCTACGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTTATGACAGCAAAAT	10989
Qy	2657	TGCTTCTGGCTCAGTTTCAATCAGTTTAAATAAATGATAAAATTTTTTGAAGCTCTGAAG	2716
Db	10990	TGCTTCTGGCTCAGTTTCAATCAGTTTAAATAAATGATAAAATTTTTTGAAGCTCTGAAG	11049
Qy	2717	ATAAAAACCAATAAAATAATATAAAGCTGATTTATGAAGTTTAAATAAATAAATCAG	2776
Db	11050	ATAAAAACCAATAAAATAATATAAAGCTGATTTATGAAGTTTAAATAAATAAATCAG	11109
Qy	2777	TATGATGGAATAAACTTGAAA	2797
Db	11110	TATGATGGAATAAACTTGAGA	11130
RESULT 4			
US-09-966-880A-15			
; Sequence 15, Application US/09966880A			
; Patent No. US2002016473A1			
; GENERAL INFORMATION:			
; APPLICANT: Honjo, Tasuku			
; APPLICANT: Muramatsu, Masamichi			
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE			
; FILE REFERENCE: 06501-088001			
; CURRENT APPLICATION NUMBER: US/09/966,880A			
; CURRENT FILING DATE: 2001-09-28			
; PRIOR APPLICATION NUMBER: PCT/JP00/01918			
; PRIOR FILING DATE: 2000-03-28			
; PRIOR APPLICATION NUMBER: JP 11-371382			
; PRIOR FILING DATE: 1999-12-27			
; PRIOR APPLICATION NUMBER: JP 11-178999			
; PRIOR FILING DATE: 1999-06-24			
; PRIOR APPLICATION NUMBER: JP 11-87192			
; PRIOR FILING DATE: 1999-03-29			
; NUMBER OF SEQ ID NOS: 36			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ. ID. NO 15			
; LENGTH: 2172			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-966-880A-15			
Query Match 77.1%; Score 2172; DB 9; Length 2172;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	623	CCCTCTGATGAGGTTTGATGACTTACGAGACGCAATTCGTACTTTTGGGACTTTGATAGCAA	682
Db	1	CCCTCTGATGAGGTTTGATGACTTACGAGACGCAATTCGTACTTTTGGGACTTTGATAGCAA	60
Qy	683	CTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTGTATAAAAAACAT	742

Db	61	CTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAACAAGT	120
Qy	743	CCTTCAAGTCCTCTCTGTTTTTATCTTCAACTCTCACTTCTTAGAGTTTACAGAAAA	802
Db	121	CCTTCAAGTCCTCTCTGTTTTTATCTTCAACTCTCACTTCTTAGAGTTTACAGAAAA	180
Qy	803	ATATTTATATAGGACTCTTTAAAAAGATCTATGCTTGAATAATAGAGAAGGACACAGT	862
Db	181	ATATTTATATAGGACTCTTTAAAAAGATCTATGCTTGAATAATAGAGAAGGACACAGT	240
Qy	863	CTGGCCAGGGACGTGCTGCAATGGTGCAGCTTTTGAATGCAACTTGTCCCCTACTGGGA	922
Db	241	CTGGCCAGGGACGTGCTGCAATGGTGCAGCTTTTGAATGCAACTTGTCCCCTACTGGGA	300
Qy	923	ATACAGAACTGACAGGACCTGGAGACCTCTAAAGTGCAAGCTTTTCTATGACTTTTA	982
Db	301	ATACAGAACTGACAGGACCTGGAGACCTCTAAAGTGCAAGCTTTTCTATGACTTTTA	360
Qy	983	GGTAGGATGAGACGAGAAGCTAGACTCTTAAAGCAATGGTGAGAGGATCAAAATGTTTTA	1042
Db	361	GGTAGGATGAGACGAGAAGCTAGACTCTTAAAGCAATGGTGAGAGGATCAAAATGTTTTA	420
Qy	1043	TATCAACATCCTTTATTTGATTGATTTGAGTTTAAAGTGGTGTGTAGTCATAGATTTT	1102
Db	421	TATCAACATCCTTTATTTGATTGATTTGAGTTTAAAGTGGTGTGTAGTCATAGATTTT	480
Qy	1103	TCATTTCTTTTCCCTTGACGTTTTCTTTCAAGTAAACAAACTCTTCCATCAGGCCATGA	1162
Db	481	TCATTTCTTTTCCCTTGACGTTTTCTTTCAAGTAAACAAACTCTTCCATCAGGCCATGA	540
Qy	1163	TCATAGGACCTCTTAATGAGATGATCTGGGTGATTTGACCCCAACCACTCTCTCCAA	1222
Db	541	TCATAGGACCTCTTAATGAGATGATCTGGGTGATTTGACCCCAACCACTCTCTCCAA	600
Qy	1223	GCATTAATATCAAACTCATGCGCTGTATGTTTTTAATCAGCAAGCACTGTTTTTATGTTG	1282
Db	601	GCATTAATATCAAACTCATGCGCTGTATGTTTTTAATCAGCAAGCACTGTTTTTATGTTG	660
Qy	1283	TACAAAGAAGATTGTTATGGTGGGGATGGAGGTATAGACCATGATGGTCACTTCAA	1342
Db	661	TACAAAGAAGATTGTTATGGTGGGGATGGAGGTATAGACCATGATGGTCACTTCAA	720
Qy	1343	GCTACTTTAATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAGACACCTTAATA	1402
Db	721	GCTACTTTAATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAGACACCTTAATA	780
Qy	1403	TGGGTGATGCTGAAGTAGCAATCTCTCGAAACGCAAACTCTTTAAGGAAGTCCCT	1462
Db	781	TGGGTGATGCTGAAGTAGCAATCTCTCGAAACGCAAACTCTTTAAGGAAGTCCCT	840
Qy	1463	AATTTAGAAACACCCACAAACTTTCATATCATATTAATTAGCAACAATTTGGAAGGAAGTTG	1522
Db	841	AATTTAGAAACACCCACAAACTTTCATATCATATTAATTAGCAACAATTTGGAAGGAAGTTG	900
Qy	1523	CTTGAATTTGGGAGAGGAAAACTTATTTGGCTCTCGTGGGCTCTTCATCTCAGAAATG	1582
Db	901	CTTGAATTTGGGAGAGGAAAACTTATTTGGCTCTCGTGGGCTCTTCATCTCAGAAATG	960
Qy	1583	CCAATCAGGTCAGGTTTGCTACATTTGTATGTGTGATGCTTCTCCCAAGGTATAT	1642
Db	961	CCAATCAGGTCAGGTTTGCTACATTTGTATGTGTGATGCTTCTCCCAAGGTATAT	1020
Qy	1643	TAACTATATAAGAGAGTTGTGACAAAACAGAAATGATTAAGCTGCGAACCGTGGCACACG	1702
Db	1021	TAACTATATAAGAGAGTTGTGACAAAACAGAAATGATTAAGCTGCGAACCGTGGCACACG	1080
Qy	1703	TCATAGTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTTGAACACAGGTGTTCAA	1762
Db	1081	TCATAGTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTTGAACACAGGTGTTCAA	1140
Qy	1763	GGCCAGCCTGGGCAACATAAACAAGATCTGTCTCTCAAAAAAATAAAAAAAGAAA	1822
Db	1141	GGCCAGCCTGGGCAACATAAACAAGATCTGTCTCTCAAAAAAATAAAAAAAGAAA	1200

RESULT 5
US-09-966-880A-1
; Sequence 1, Application US/09966880A

; Sequence 1, Application US/09966880A

Patent No. US20020164743A1
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966, 880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR APPLICATION NUMBER: JP 11-87192
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2440
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (93)...(686)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)...(92)
FEATURE:
NAME/KEY: 3'UTR
LOCATION: (690)...(2440)
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(2440)
OTHER INFORMATION: n = A,T,C or G
US-09-966-880a-1

Query Match 21.4%; Score 603.4; DB 9; Length 2440;
Best Local Similarity 69.4%; Pred. No. 2e-120;
Matches 1006; Conservative 0; Mismatches 351; Indels 92; Gaps 10;
QY 4 GAACCATCATTAATGAAGTGTAGATTTTCTGGCCTGTGAGACTTTCGAGGGAGGCAAG-AAG 62
DB 16 GAAGCAGCTTGTCTTGAAGCAAGCTTCTTGGCCTTAAGACTTTGAGGGAGTCAAGAAAG 75
QY 63 ACACCTGTGACACCATATGGACAGCTCTTGTATGAACGGGAGGAGTTTCTTTACCAAT 122
DB 76 TCACGCTGGAGACCGATATGGACAGCTTCTGTATGAAGCAAGAAAGATTTCTTTTACCATT 135
QY 123 TCAAAAATGTCCGCTGGGCTAAGGTCGGCGTGTGACACCTACCTGTGTAGTGTAGTGAAGA 182
DB 136 TCAAAAATGTCCGCTGGGCTAAGGTCGGCGTGTGACACCTACCTGTGTAGTGTAGTGAAGA 195
QY 183 GCGGTGACAGTGTACATCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGGCT 242
DB 196 GGAGAGATATGCCACCTCTGCTCACTGGACTTTCGGCCACCTTCGCAACAAGTCTGGCT 255
QY 243 GCCAGTGAATGTCTTCTCCGCTACATCTCGGACTTGGGACCTTAGACCTTGGCCGCT 302
DB 256 GCCAGTGAATGTGTCTTCTACGCTACATCTCGACACTGGGACCTGGGACCCGGCGGCT 315
QY 303 GCTACCGCTACCTCGTTTCACTCCTGTGGAGCCCTGTGTACGACTGTGCCCGACATGTGG 362
DB 316 GTTACCGCTACCTCGTTTCACTCCTGTGGAGCCCTGTGTACGACTGTGCCCGACATGTGG 375
QY 363 CCGACTTTTCTGGAGGGAACCCCACTCAGTCTGAGGATCTTTCACCGCGCCTTACT 422
DB 376 CTGAGTTTCTGAGTGGAAACCTTAACCTCAGCTGAGGATTTTTCACCGCGCCTTACT 435
QY 423 TCTGTGAGACCGCAGGCTGAGCCGAGGGCTGGCGGCTGCGACCGCGCGGGTGC 482
DB 436 TCTGTGAGACCGCAGGCTGAGCCGAGGGCTGCGGAGACTGTCACCGCGCTGGGGTCC 495

RESULT 6
US-09-764-891-5477
; Sequence 5477, Application US/09764891
; Publication No. US2003007780A1

QY 483 AAATGACCATCATGACCTTCAAGATTAATTTTACTGCTGAATACATTTTGTAGAAAACC 542
DB 496 AGATCGGATCATGACCTTCAAGACTATTTTACTGCTGAATACATTTTGTAGAAAACC 555
QY 543 ATGAAAGAACTTTCAAGACCTTGGGAGGCTGATGAAATTTAGTTGCTCTCTCCAGAC 602
DB 556 GTGAAAGAACTTTCAAGACCTTGGGAGGCTGATGAAATTTCTGTCGGCTTAACACGAC 615
QY 603 AGCTTCGGCGCATCTTTTCCCTGTATGAGGTTGATGCTTACGAGACCATTTCTGTA 662
DB 616 AACTTCGGCGCATCTTTTCCCTGTATGAGGTTGATGCTTACGAGACCATTTCTGTA 675
QY 663 CTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTGAA 722
DB 676 TGTGGGATTTTGAAGCAACCTCTCTGGAATGTACACGATGAAATTTCT---CTGAA 732
QY 723 GACAGTGGATAAAGAACAGCTCTCTCAAGTCTCTCTCTTTTATTTATCTTCAACTCTCACT 782
DB 733 GAGACTGGATAGAAAACAAACCTTCA---ACTACATGTTTTTCTTCTTAAGTACTCACT 790
QY 783 TCTTAGAGTTTACAGAAAATAATTTATATACGACTCTTTTAAAGATCTATGCTTGTAA 842
DB 791 TTATAGTGTAGGGGAAA-----TTATATGACTTTTAAAGATCTATGCTTGTAA 843
QY 843 AATAGAGAAGGAACACAGGCTTGGCCAGGACGCTGCTGCAATTTGGTGCAGTTTGAATGC 902
DB 844 AGGACCCAGAGCAATGATTAAC-----TGAGCTTGTCTGTGC 882
QY 903 AACATTGTCCCTTACT---GGGAATAACAGAACTGCGAGGACCTGGGAGGATCTCTAAAGTGT 961
DB 883 AACATCGCATCTACTTGGGAACAGCAATAACTTCCAGACTTTGGGCTGCTGAATGATGCTC 942
QY 962 AACGTTTTCTATGACTTTTAGTGTAGGATGAGACCAAGGTAGATCTCTAAAGACATGG 1021
DB 943 TTTTCTTCAACAGCAT-----GGAAAAGCATAT 971
QY 1022 TGAGAGGATCAAAATGTTTTATATCAACATCCTTTATTTATTTGATTTGATTTGAGTTAACA 1081
DB 972 GGAGACGACACACAGTTTGTACACCCACCTGTGTCTTGTATTTGATTTGATTTGATTTGAT 1031
QY 1082 GTGGTGTAGTGTAGATTTTCTATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1141
DB 1032 GGGGTATCAGTACGAGGATTTCTTATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1084
QY 1142 AACTCTTCCATCAGGCAATGATCTATAGGACCTCTATAGAGATCTCTGGGTGATTTGTG 1201
DB 1085 ---GGTCTTTTCTGACAGGTCACGGGCTGCTCTACAGTCTCTGCTGAGCAAT----- 1137
QY 1202 ACCCAACCATCTCTCAAGCATTTAATATCAATCATGCTGCTGTATGTTTAAATCAAGC 1261
DB 1138 -CACAGCCATTTCTCTCAAAAACATTAATACTCAGGCACATGCTGTATGTTTCA----- 1191
QY 1262 AGAAGCATGTTTTATGTTTTGTACAAAGAAAGATTTGTTATGGGTGGGATGGAGGTATAG 1321
DB 1192 -----CTGTCCGCTGCTGTTTTTTCACATTTGATGTGAAGGGCTTGGGGTGGATTTGA 1245
QY 1322 ACCATGATGTCACCTTCAAGCTACTTTTAAAGATCTTTAAATGGGAGGAGGACT 1381
DB 1246 AGAATGACCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
QY 1382 GTGAACAGACACCTTAATATGGTTGATGCTGTGAAGTGAAGTAAATCTTCTGGAAACGCA 1441
DB 1306 ACAGAAGAACTACTCTGAAATGAGTTTCAAGCCCTCAAGAACCAATCCCTTGGAAACACA 1365
QY 1442 AACTCTTTT 1450
DB 1366 GACTCTTTT 1374


```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5477
; LENGTH: 13862
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-891-5477

Query Match      10.5%; Score 295.2; DB 9; Length 13862;
Best Local Similarity 74.6%; Pred. No. 3e-53;
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1667 AACAGAATGATAAGCTGCGAACCGTGGCACACGCTCATAGTTCCTAGCTGTGGGAGG 1726
DB 1508 AAACCTTAATAAGCTGGCTGCTGTGGCTCATGCTGTGATCCCGACACTTTGGGAGG 1567
QY 1727 TTGAGGAGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATACAAG 1786
DB 1568 CTGAGGAGGAGGATCACTTTGAGGTGAGGAGTTTGAGACGAGCCTGGCCAAACATGGCGAA 1627
QY 1787 ATCCTGTCTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
DB 1628 ACCCATCTCTACTAAATAATACAAAAGTTAGCCAGGCATGGGCGGCGCATGGTGGTGGC 1687
QY 1846 TCACGCTGTAAATCCAGCACTTTGGGAGCGGAGCGGCGGCGGATCACCTGTGGTCAGGA 1905
DB 1688 TCATGCTGTAAATCCAGCACTTTGGGAGCGGAGGTTGGCGGATCACCTGAGGTCAGGA 1747
QY 1906 GTTTGAGACCGCCTGCCACATGCGAATGCGAATGCGAATGCGAATGCGAATGCGAATGCGA 1965
DB 1748 GTTCAAGACCGCCTGGCAACATGCGAATGCGAATGCGAATGCGAATGCGAATGCGAATGCGA 1807
QY 1966 GCCAGCGTGGTAGCAGGACCTGTAAATCCAGCTACTTTGGGAGGCTGAGGCGAGGAGAAAT 2025
DB 1808 GCCAGCTTTATGTTAGGCGCCTATATCCAGCTACTCGGAAGGCTGAGGCGAGGAGAAAT 1867
QY 2026 CGTTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGGCTTGCACCTCCAGCC 2085
DB 1868 CGTTTGAACCCAGGAGGCGGAGTTTGCACCGAGCGGAGATTTGCCACTGCACCTCCAGCC 1927
QY 2086 TGGGCGACAGAGCAAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2145
DB 1928 TGGGCGATAA-AGCGAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
QY 2146 AGAACAATATTTGGGAGAGAGAGGATGGGAAGCATTTGCAAGGAAATTTGTG 2197
DB 1987 AGAAAAATTAGCCAGCGTGGTGGCATGCACCTGTAGTCTAGCTACTTTGGG 2038

RESULT 7
US-09-764-891-10204
; Sequence 10204, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10204
; LENGTH: 13862
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match      10.5%; Score 295.2; DB 9; Length 13862;
Best Local Similarity 74.6%; Pred. No. 3e-53;
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1667 AACAGAATGATAAGCTGCGAACCGTGGCACACGCTCATAGTTCCTAGCTGTGGGAGG 1726
DB 1508 AAACCTTAATAAGCTGGCTGCTGTGGCTCATGCTGTGATCCCGACACTTTGGGAGG 1567
QY 1727 TTGAGGAGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATACAAG 1786
DB 1568 CTGAGGAGGAGGATCACTTTGAGGTGAGGAGTTTGAGACGAGCCTGGCCAAACATGGCGAA 1627
QY 1787 ATCCTGTCTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
DB 1628 ACCCATCTCTACTAAATAATACAAAAGTTAGCCAGGCATGGGCGGCGCATGGTGGTGGC 1687
QY 1846 TCACGCTGTAAATCCAGCACTTTGGGAGCGGAGCGGCGGCGGATCACCTGTGGTCAGGA 1905
DB 1688 TCATGCTGTAAATCCAGCACTTTGGGAGCGGAGGTTGGCGGATCACCTGAGGTCAGGA 1747
QY 1906 GTTTGAGACCGCCTGCCACATGCGAATGCGAATGCGAATGCGAATGCGAATGCGAATGCGA 1965
DB 1748 GTTCAAGACCGCCTGGCAACATGCGAATGCGAATGCGAATGCGAATGCGAATGCGAATGCGA 1807
QY 1966 GCCAGCGTGGTAGCAGGACCTGTAAATCCAGCTACTTTGGGAGGCTGAGGCGAGGAGAAAT 2025
DB 1808 GCCAGCTTTATGTTAGGCGCCTATATCCAGCTACTCGGAAGGCTGAGGCGAGGAGAAAT 1867
QY 2026 CGTTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGGCTTGCACCTCCAGCC 2085
DB 1868 CGTTTGAACCCAGGAGGCGGAGTTTGCACCGAGCGGAGATTTGCCACTGCACCTCCAGCC 1927
QY 2086 TGGGCGACAGAGCAAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2145
DB 1928 TGGGCGATAA-AGCGAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
QY 2146 AGAACAATATTTGGGAGAGAGAGGATGGGAAGCATTTGCAAGGAAATTTGTG 2197
DB 1987 AGAAAAATTAGCCAGCGTGGTGGCATGCACCTGTAGTCTAGCTACTTTGGG 2038

RESULT 8
US-10-205-428-1003
; Sequence 1003, Application US/10205428
; Publication No. US20030108907A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PALL7C1
; CURRENT APPLICATION NUMBER: US/10/205,428
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
```



```

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 15297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-003-295-3

```

Query Match	9.9%;	Score 280;	DB 9;	Length 15297;
Best Local Similarity	74.9%;	Pred. No. 6.3e-50;		
Matches 370;	Conservative 0;	Mismatches 110;	Indels 14;	Gaps 1;
QY	1641	ATTAACATATATAGAGAGCTTGTCACAAAACAGAAATGATAAAGCTCGGAACCGTGGCACAC	1700	
Db	11651	ACTCCCTTAATGCCAACCTTCCACGAGCAGAAATGAAGATAAATCTGCCAGTTGCTCAC	11710	
QY	1701	GCTCATAGTCTTAGCTTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGTGTTTC	1760	
Db	11711	GCCTGTCATCCAGCACCTTGGGAGGCTGAGCTGGGTGGATCACTTGAGCCCGAGGATTC	11770	
QY	1761	AAGCGCAGGCTGGGCAACATACAGATCTCTCTCAAAAAAATAAAAAAAGA	1820	
Db	11771	AAGATCAGCTTGGACACACAGTGAACCTCCATCTGTACAAAAATACAAAAA	11825	
QY	1821	AAGAGAGAGCGCGGCGTGGTGCTCAGGCTGTAAATCCAGCACTTTGGGAGGCCGAG	1880	
Db	11826	-----GACTGGGCAGGTTGGCTCACACCTGTAAATCCAGCACTTTGGGAGGCCGAG	11876	
QY	1881	CGGGCGGATCACTCTGCTCAGGAGTTTGAGACCAGGCTGGCCAACTATGCAAAACCCC	1940	
Db	11877	GCAGGTGGATCACTCTGGTCAGGAGTTTGAGACCAGCCAGCAACATGGTGAACCCC	11936	
QY	1941	GTCTGTACTCAAAATGCAAAATTAGCCAGGCGTGGTAGCAGGCACCTGTATCCAGCT	2000	
Db	11937	ATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGTGCCTGTAAATCCAGCT	11996	
QY	2001	ACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGTTTCAGTAAGCT	2060	
Db	11997	ACTTGGGAGGCTGAGGTGGGAGAAATGGTTGAACCCAGGAGCGAGGCTGCAGTGAACC	12056	
QY	2061	GAGATGTCGCCCTTGCACTCCAGCCTGGCGCAGCAAGACAGACTCTCTCTCAGAAAAA	2120	
Db	12057	GAGATGTGCCACTGCATCCAGCCTGGCGCAGCAAGAGTGAATCCATCTCAAAAAA	12116	
QY	2121	AAAAAAAAGAGA	2134	
Db	12117	CCAAAAACAAAAA	12130	

Search completed: June 19, 2003, 11:33:58
Job time : 381.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:09:48 ; Search time 115.983 Seconds
(without alignments)
6475.074 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 4948
Sequence: 1 adagacaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1815940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=Cpn2_1/USPTO.spool/US09966880/runat_14062003_175523_10297/app_query.fasta_1.9493
-DB=A_Geneseq_101002 -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US09966880 @CGN_1.1.632 @runat_14062003_175523_10297 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1086	21.9	198	21	AA24198	Human activation-1
2	1008	20.4	198	21	AA24197	Mouse activation-1
C 3	417	8.4	188	22	ABG11932	Novel human diagno
C 4	402	8.1	126	22	AB11478	Human secreted pro
C 5	394	8.0	110	21	AA43292	Human ORFX ORF3056
6	390	7.9	384	20	AA42383	Amino acid sequenc
7	390	7.9	384	22	AA39075	Human secreted pro
8	390	7.9	384	23	AB55784	Human polypeptide
9	388	7.8	119	22	AA83429	Human immune/haema
10	388	7.8	384	21	AA84437	Amino acid sequenc
C 11	386	7.8	107	22	AA31902	Novel human secret
C 12	383.5	7.8	134	22	AB12093	Human secreted pro
C 13	381.5	7.7	239	22	AB90850	Human secreted pro
C 14	381.5	7.7	310	22	AB90647	Human secreted pro
C 15	378	7.7	165	21	AA42917	Human ORFX ORF2681
C 16	377	7.6	381	22	AA30235	Novel human secret
C 17	376	7.6	175	22	AA31725	Novel human secret
C 18	375.5	7.6	124	23	AA80385	Human neurofibroma
C 19	375.5	7.6	382	22	AA31818	Novel human secret
C 20	375.5	7.6	382	22	AA32610	Novel human secret
C 21	375.5	7.6	382	22	AA32707	Novel human secret
C 22	371.5	7.5	121	22	AA32891	Novel human secret
C 23	369.5	7.5	190	19	AA77092	Human RNA editing
24	369.5	7.5	190	21	AA11373	Human RNA editing
25	369.5	7.5	190	22	AA38905	Human polypeptide
26	369.5	7.5	210	22	AA23550	Novel human enzyme
27	369.5	7.5	221	22	AA40691	Human polypeptide
28	369.5	7.5	222	21	AA57061	Human prostate can
29	369.5	7.5	222	22	AA23537	Novel human enzyme
C 30	367.5	7.4	153	22	AA31783	Novel human secret
C 31	366.5	7.4	116	22	AA47436	Novel bone marrow
C 32	365	7.4	94	23	ABJ03693	Human ovary specif
C 33	364	7.4	209	22	AA012667	Human polypeptide
C 34	363	7.4	99	22	AB11839	Human novel protei
C 35	361	7.3	168	22	ABG0628	Novel human diagno
C 36	360.5	7.3	127	23	ABG64878	Human albumin fusi
C 37	360.5	7.3	127	23	ABG20800	Human gene 8 encod
C 38	359	7.3	144	23	AB79196	Human prostate spe
C 39	358.5	7.3	134	23	AB21703	Human secreted pro
C 40	358	7.3	100	23	ABP51849	Human colon specif
C 41	358	7.3	126	22	AA64466	Human gas vesicle
C 42	355	7.2	101	22	ABG01434	Novel human diagno
C 43	353.5	7.2	218	22	AAE11963	Human c-myb-relate
C 44	353	7.2	166	22	AA32028	Novel human secret
C 45	352	7.1	146	22	AB11580	Human secreted pro

ALIGNMENTS

RESULT 1
AA24198
ID AA24198 standard; Protein; 198 AA.
XX
AC AA24198;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human activation-induced cytidine deaminase SEQ ID NO:8.
XX
KW Activation-induced cytidine deaminase; AID: cytidine deaminase;
KW Immune related disease; allergy; allergic disease; anti-allergic;
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; IgG subclass selection disorder.
 OS Homo sapiens.
 XX WO200058480-A1.
 PN 05-OCT-2000.
 XX 28-MAR-2000; 2000WO-JP01918.
 XX 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX (NIBS) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX Honjo T, Muramatsu M;
 PI WPI; 2000-611715/58.
 XX N-PSDB; AAC55312.
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX Claim 1: Page 140-141; 174pp; Japanese.
 XX The present sequence is human activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antisthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX Sequence 198 AA;
 XX
 Alignment Scores:
 Pred. No.: 5.3le-114 Length: 198
 Score: 1086.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 21.95% Indels: 0
 DB: 21 Gaps: 0
 US-09-966-880a-7 (1-2818) x AAB24198 (1-198)
 QY 80 ATGACACGCTCTTGATGAACCGGAGGAGCTTTCTTACCAATTCAGAGGCGTGG 139
 DB 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
 QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGCTACA 199
 DB 21 AlaLysGlyArgArgGluThrTyrLeuLeuTyrValValLysArgArgAspSerAlaThr 40
 QY 200 TCCTTTTCTACGTGACCTTGGTTATCTTCGCAATTAAGAACGGCTGCCAGTGGATGCTC 259
 DB 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 QY 260 TTCTCCGCTACATCTCGGACTGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG 319
 DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

QY 320 TTCACCTCTCTGGAGCCCTCTACGACTGTGTCCCGACATGTGCCGACTTTCTGCGAGGG 379
 DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 439
 DB 101 AsnProAsnLeuSerLeuArgGlyPheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 QY 440 GCTGAGCCGAGGCGTGGCGGCTGCACCGCGCGGCTGCAATAGGCATCATGACC 499
 DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 QY 500 TTCAAAGATTATTTTACTCTGGAATACTTTTGTAGAAACCATGAAGAACTTTCAAA 559
 DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 QY 560 GCCTGGAGGCGTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 619
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 QY 620 TTGCCCTCTGATGAGTTGATGACTTACGAGACGCAATTTCTGCTACTTTGGGACTT 673
 DB 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 RESULT 2
 AAB24197
 ID AAB24197 standard; Protein; 198 AA.
 XX AAB24197;
 AC AAB24197;
 XX 05-FEB-2001 (first entry)
 DT Mouse activation-induced cytidine deaminase SEQ ID NO:2.
 XX
 DE Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW Immune related disease; allergy; allergic disease; antiallergic;
 KW antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disease;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder.
 XX Mus musculus.
 OS
 XX WO200058480-A1.
 PN 05-OCT-2000.
 XX 28-MAR-2000; 2000WO-JP01918.
 XX 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX (NIBS) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX Honjo T, Muramatsu M;
 PI WPI; 2000-611715/58.
 XX N-PSDB; AAC55307.
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX Claim 1: Page 131-132; 174pp; Japanese.
 PS The present sequence is mouse activation-induced cytidine deaminase
 XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC

CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class)
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX
 SQ Sequence 198 AA;

Alignment Scores:
 Pred. No.: 3,99e-105 Length: 198
 Score: 1008.00 Matches: 183
 Percent Similarity: 95.94% Conservative: 6
 Best Local Similarity: 92.89% Mismatches: 8
 Query Match: 20.37% Indels: 0
 DB: 21 Gaps: 0

US-09-966-880A-7 (1-2818) x AAB24197 (1-198)

QY 80 ATGCAGAGCTCTGTGATGAACCGAGAGAGTCTTTTACCAATTCAAAATGCCGCTGG 139
 Db 1 MetAspSerLeuLeuMetLysGlnLysPheLeuTyrHisPheLysAsnValArgTrp 20
 QY 140 GCTAAGGGTGGCGCTGAGACCTACCTGTCTGTCTAGTGAAGGCGTACAGTGTACA 199
 Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
 QY 200 TCTTTTCTAGTACCTTTGGTTATCTTCGCAATAAGACGGCTGCCAGCTGGAAATTCCTC 259
 Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeuLeu 60
 QY 260 TTCCTCGCTACATCTCGAGCTGGACCTAGACCTCGCGCTGCTACCGCTGCTACCTCG 319
 Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 QY 320 TTCACCTCTGGAGCCCTCTCTACGACTGTGCCGCTGCTGCGGCTCTCTCTGCGAGGG 379
 Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
 QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTCTGTGTAGGACCGCAAG 439
 Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 QY 440 GCTGAGCCGCGGCTGGCGGCTGCACCGCGCGGCTGCAAAATAGCATCATGACC 499
 Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140
 QY 500 TTCAAAGATATTTTACTGCTGGAATCTTTTGTAGAAACCATGAAGAACTTTCAAA 559
 Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160
 QY 560 GCCTGGAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGCATCCTT 619
 Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180
 QY 620 TTCGCCCTGATCAGCTGATGACTTACGAGAGCGCATTTCTGTTTGGGA 670
 Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3

ID ABG11932

XX standard; Protein; 188 AA.

AC ABG11932;

XX

DT 18-FEB-2002 (first entry)
 XX Novel human diagnostic protein #11923.
 DE
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PR
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS76119.
 DR
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 20; SEQ ID No 42291; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 188 AA;

Alignment Scores:
 Pred. No.: 6,86e-38 Length: 188
 Score: 417.00 Matches: 99
 Percent Similarity: 64.40% Conservative: 24
 Best Local Similarity: 51.83% Mismatches: 41
 Query Match: 8.45% Indels: 28
 DB: 22 Gaps: 4

US-09-966-880A-7 (1-2818) x ABG11932 (1-188)

QY 2126 TTTTCTTCTGAGACAGAGTCTTGTCTTGTCCGCCAGCTGGAGTGCACGGCAC 2067
 Db 1 PheAsnPhePheLeuArgGlnSerLeuAlaLeuLeu-ProArgLeuGluCysSerGlyMe 20
 QY 2066 GATCTCAGCTTACTGCAACCTCCACCTCTGGGTTCAGCGATTCTCTCCCTCAGCCTC 2007
 Db 20 ttleSerAlaHisCysAsnLeuHisLeuProGlySerSerAsnSerProAlaLeuAlaPr 40

QY 2006 CCAAGTAGCTGGGATTACAGGTGCTGCTACACGCCCTGGCTAAATTTTTCGATTTTGGT 1947
Db :|||||
40 oargValAlaGlyIlelleSerThrCysHisAlaGlnValIlePheAlaCysLeuVa 60
QY 1946 ACAGACGGGTTTTCCTGCTGGCCAGGCTGCTCAACCTCTGACACAGGTGATPCC 1887
Db :|||||
60 loMetGlyPheHsHsValGlyGlnAlaGlyLeuLysLeuThrSerGlyAspLe 80
QY 1886 GCCCGCTCGGCTCCCA-AAAGTGTGGGATTACAGGCTGAGCCACACCGCCGCT 1828
Db :|||||
80 uProAlaLeuAlaSerGluLysCysTrpAspTyrArgGluGlnProArgProAlaLe 100
QY 1827 CTCCTCTCTCTTTTTCCTGAGACAGGATCTTGTATGTTGTCGCCAGGC 1768
Db :|||||
100 u-----PhePheTrpTyrPheSerArgAspGlnArgPheThrLeuGlyGlnAl 117
QY 1767 TGGCTTCAACACCTGTGTTCAAGCCATCTCTCCCTCAACCTCCCAAGCAGCTAGAAC 1708
Db :|||||
117 aGlyLeuGluLeuLeuAlaSerSerAsnProProAlaSerAlaSerLysSerAlaGlyI 137
QY 1707 TATGAGCTGTGCCAGGCTTCG-----CAGCTTATCATCTGTT----- 1668
Db :|||||
137 eThrAspValSerHisGlyThrHisAlaLysArgLeuTyr-SerTrpLeuCysValSerG 157
QY 1667 -----TTGTCAACACTCTCTTATAGTTAA 1642
Db :|||||
157 uThrAlaPheAsnProAsnPheSerThrPheArgLeuLysAsnIleLeuPheMetValas 177
QY 1641 TATACCTTTGGGAGAGCATCATACATAC 1611
Db :|||||
177 nLeu-----GluHisAsnTyr 182

RESULT 4
ABBI1478
ID ABBI1478 standard; peptide; 126 AA.
AC ABBI1478;
XX
XX
XX 11-JAN-2002 (first entry)
XX
XX Human secreted protein homologue, SEQ ID NO:1848.
XX
XX Human; cytokine; cell proliferation; tissue growth; immunomodulation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytosolic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.
XX
OS Homo sapiens.
XX
XX W0200157188-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457740/49.

N-PSDB; ABA08722.
Human proteins and DNA encoding sequences useful for preventing,
treating or ameliorating a medical condition in a mammalian subject
e.g. arthritis and cancer -
Claim 20; Page 197; 1963pp; English.
Sequences ABBI0981-ABBI12330 represent 1350 novel human polypeptides, and
sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
invention also relates to vectors and recombinant host cells comprising a
nucleotide of the invention, methods of producing the novel polypeptides,
antibodies against the polypeptides, methods of detecting the nucleotides
or polypeptides in a sample, and methods of identifying compounds which
bind to polypeptides of the invention. Although novel, many of the
polypeptides of the invention have homology to known proteins, and hence
giving an insight into their probable biological activities, and hence
potential therapeutic applications. The polypeptides of the invention may
have various activities, including cytokine, cell proliferation or cell
differentiation activities; stem cell growth factor activity;
haematopoiesis regulatory activity; tissue growth activity;
immunomodulatory activity; activin- or inhibin-related activities;
chemotactic or chemokinetic activities; haemostatic, thrombotic or
thrombolytic activities; receptor or ligand activities; or may be
involved in oncogenesis, cancer cell proliferation or metastasis.
The invention on their biological activities, polypeptides and nucleotides of
conditions, e.g., by protein or gene therapy. Such conditions include
cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
proliferative retinopathy, atherosclerosis, coronary heart disease,
arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
vascular growth. Polypeptides involved with tissue regeneration and
repair (or nucleic acids encoding them) may be used to promote wound
healing (e.g., of burns, incisions and ulcers), while those with
immunomodulatory activities may be used in the treatment of viral,
bacterial and fungal infections in addition to immune disorders.
Polypeptides with growth factor activity may be used in cell cultures to
promote cell growth. For example, such polypeptides may be used to
manipulate stem cells in culture to give rise to neuroepithelial cells
that can be used to augment or replace cells damaged by illness,
autoimmune disease or accidental damage. The polypeptides and nucleotides
may also be used in the diagnosis of the above conditions, and in drug
screening techniques. The present sequence represents a novel human
polypeptide of the invention.
SQ Sequence 126 AA;
Alignment Scores:
Pred. NO.: 2.77e-36 Length: 126
Score: 402.00 Matches: 90
Percent Similarity: 72.39% Conservative: 7
Best Local Similarity: 67.16% Mismatches: 22
Query Match: 8.14% Indels: 15
DB: 22 Gaps: 3
US-09-966-880A-7 (1-2818) x ABBI1478 (1-126)
QY 2183 AATGCTTCCCATCTCTCTCT-----CCCAATATGTTCTCTCTCTCTCTCT 2136
Db :|||||
2 AsnAlaSerThrValTyrSerSerGlnGlyAspProLys----- 14
QY 2135 CTCCTCTTTTTCCTTTTTCCTGAGACAGAGTCTGCTGTCGCCAGCGTGAGTG 2076
Db :|||||
15 -----SerPhePheLeuLeuArgTrpSerLeuAlaLeuValAlaGlnAlaGlyGlu 32
QY 2075 CAACGGCAGCATCTCAGCTTACTGCAACCTCCACCTCTGGTTCAGCGATTCTCTGC 2016
Db :|||||
33 Gln***ArgAspLeuSerSerLeuGlnProProProGlyPheLys***PheSerCys 52
QY 2015 CTCAGCCTCCCAAGTAGCTGGGATTACAGGTGCTGCTACAGCCCTCTTAATTTTTC 1956
Db :|||||
53 LeuSerLeuProSerSerTrpAspTyrArgCysProLeuProCysLeuAlaAsn-Phe** 72

CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
 CC are deposited with the American Type Culture Collection (ATCC) with
 CC accession number 98599. The polynucleotides and encoded polypeptides have
 CC cytotatic, anti-inflammatory, immunomodulator, vulnerary,
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic
 CC and anti-inflammatory activity and acting as cytokine modulators,
 CC haematopoiesis regulators, tissue growth modulators and/or cadherin
 CC suppressors. The polypeptides and polynucleotides are useful in gene
 CC therapies, particularly for preventing, treating or ameliorating any of
 CC the following diseases: immune deficiency and disorders; e.g. bacterial
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or
 CC osteoarthritis; central and peripheral nervous system diseases and
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or
 CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
 CC foliaceus.
 XX
 SQ Sequence 384 AA;

Alignment Scores:
 Pred. No.: 1.23e-34 Length: 384
 Score: 390.00 Matches: 83
 Percent Similarity: 61.62% Conservative: 31
 Best Local Similarity: 44.86% Mismatches: 59
 Query Match: 7.88% Indels: 12
 DB: 23 Gaps: 3

US-09-966-880A-7 (1-2818) x ABB55784 (1-384)

QY 95 ATGAACCGGAGGAGTTCTTTACCAATTCARAAATTCGCTGGCTAAGGTCGGCGT 154
 Db 197 MetaspProthrPheThrPheAsnPheAsnGluProtrpValArgGlyArgHis 216
 QY 155 GAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCTTTTTCACATGGAC 214
 Db 217 GluThrTyLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236
 QY 215 -----TTTGGTTATCTTCCGTAATAAG-----AACGGCTGC 244
 Db 237 GlnArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256
 QY 245 CAGTGTGAATGTCTTCTCGCTACATCTCGACTGGGACCTAGACCTGCGCGCTGC 304
 Db 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspGlnAsp 276
 QY 305 TACCGCGTCACCTGGTTACCTCTCGAGCCCTGCTACGACTGTGCGGACATGTGGCC 364
 Db 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296
 QY 365 GACTTCTCGAGGAGACCCACCTCAGTCTGAGGATCTTACCGCGGCTCTCTACTTC 424
 Db 297 LysPheIleSerLysAsnLysHisValSerLeuLysPheThrAlaArgIleTyr--- 315
 QY 425 TGTGAGGACCGCAAGGCTGAGCGGCTGCGCGCTGCGACCGCGCGGGTCA 484
 Db 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaLys 334
 QY 485 ATAGCATCATGACCTTCAAGATATTTTACTGCTGGAATCTTTTGTAGAAACCAT 544
 Db 335 IleSerIleMetThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354
 QY 545 GAAGAACCTTCAAGCCCTGGAAGGCTGCATGAAATCTAGTCTCTCTCCAGACAG 604
 Db 355 GlyCysProPheGlnProtrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374
 QY 605 CTTGGCGCATCTT 619
 Db 375 LeuArgAlaIleLeu 379

RESULT 9
 AAM83429
 ID AAM83429 standard; Protein; 119 AA.
 XX
 AC AAM83429;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human Immune/haematopoietic antigen SEQ ID NO:11022.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytotatic; gene therapy; vaccine; metastasis.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180828.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0225119.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235836.
 PR 27-SEP-2000; 2000US-0235837.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246612.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 DR N-PSDB; AAK56210.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Claim 11; SEQ ID NO 11022; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 119 AA;
 Alignment Scores:
 Pred. No.: 1.05e-34 Length: 119
 Score: 388.00 Matches: 78
 Percent Similarity: 82.00% Conservative: 4
 Best Local Similarity: 78.00% Mismatches: 17
 Query Match: 7.84% Indels: 1
 DB: 22 Gaps: 0
 US-09-966-880A-7 (1-2818) x AAM83429 (1-119)
 QY 1832 CCGGGCGTGGTGCAGCGCTGTAATCCAGCACCTTTGGAGGCGGAGCGCGGATC 1891
 Db 21 ProGlySerVal-LeuThrProValIleProThrLeuTrpGluAlaGluAlaGlyGlySe 40
 QY 1892 ACCTGTGGTGGAGGATTTGAGACAGCGCTGGCCAAACATGGCAAAACCCGCTGTACTCA 1951
 Db 40 rProGluValArgSerArgProAlaTrpProThrTrpGlyAsnLeuPheSerThrLy 60
 QY 1952 AAATGCAAAATTTAGCCAGCGGTGGTAGCAGGACCTGTATCCAGCTACTTGGGAGGC 2011
 Db 60 sAsnThrLysIleSerArgAlaTrpTrpGlnLeuProValIleProAlaThrGlnGluAl 80
 QY 2012 TGAGGCGAGGAGATCGCTTGAACCCAGGAGGTTCAGTTCAGTAAGTCGATCGTCC 2071
 Db 80 aGluAlaGlyGluSerLeuGluProArgTrpArgLeuGlnTrpAlaLysValAlaPr 100
 QY 2072 GTTGCACTCCAGCGCTGGCGACAGAGCAAGACTCTGCTCAGAAAAAAGAAAAA 2129

Db 100 oLeuHisSerLeuGlySerLysSerLysThrSerSerGlnLysLysLysLys 119
 RESULT 10
 AAY84437
 ID AAY84437 standard; Protein; 384 AA.
 AC AAY84437;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a human RNA-associated protein.
 XX
 KW Human; RNA-associated protein; cell proliferation; cancer; inflammation;
 KW immune response; reproductive disorder; actinic keratosis;
 KW atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;
 KW mixed connective tissue disease; myelofibrosis; primary thrombocytopenia;
 KW paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;
 KW trauma.
 KW Homo sapiens.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT Modified-site 13
 FT /note= "potential phosphorylation site"
 FT Modified-site 19
 FT /note= "potential phosphorylation site"
 FT Modified-site 28
 FT /note= "potential phosphorylation site"
 FT Modified-site 32
 FT /note= "potential phosphorylation site"
 FT Modified-site 86
 FT /note= "potential phosphorylation site"
 FT Modified-site 148
 FT /note= "potential phosphorylation site"
 FT Modified-site 167
 FT /note= "potential phosphorylation site"
 FT Modified-site 277
 FT /note= "potential phosphorylation site"
 FT Modified-site 311
 FT /note= "potential phosphorylation site"
 FT Modified-site 327
 FT /note= "potential phosphorylation site"
 FT Modified-site 349
 FT /note= "potential phosphorylation site"
 FT Modified-site 372
 FT /note= "potential phosphorylation site"
 FT Modified-site 427
 FT /note= "potential phosphorylation site"
 XX
 PN WO200015799-A2.
 XX
 PD 23-MAR-2000.
 XX
 XX 17-SEP-1999; 99WO-US21688.
 XX
 XX 17-SEP-1998; 98US-0156039.
 PR 22-SEP-1998; 98US-0158720.
 PR 04-NOV-1998; 98US-0186815.
 PR 08-APR-1999; 99US-0128660.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;
 PI Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J;
 XX
 DR WPI; 2000-271437/23.
 DR N-PSDB; AAA12409.
 XX
 XX New polypeptides and polynucleotides, useful for preventing and
 PT treating a disorder associated with increased or decreased expression
 PT of RNA associated proteins -
 XX

PS Claim 1; Page 101-102; 131pp; English.
 XX
 CC The present sequence represents a human RNA-associated protein. The
 CC expression of RNA-associated proteins is closely associated with
 CC reproductive tissues, nervous tissues, cell proliferation including
 CC cancer, inflammation and immune responses, and so they may be used
 CC for diagnosis, treatment or prevention of cell proliferative,
 CC immune/inflammatory disorders, and reproductive disorders. Diseases
 CC and disorders which may be treated include actinic keratosis,
 CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis,
 CC mixed connective tissue disease, myelofibrosis, paroxysmal nocturnal
 CC hemoglobinuria, polycythemia vera, psoriasis, primary thrombocytopenia
 CC and cancers, and trauma.
 XX
 SQ Sequence 384 AA;
 Alignment Scores: 2.07e-34 Length: 384
 Pred. No.: 388.00 Matches: 82
 Score: 61.62% Conservative: 32
 Best Local Similarity: 44.32% Mismatches: 59
 Query Match: 7.84% Indels: 12
 DB: 21 Gaps: 3
 US-09-966-880A-7 (1-2818) x AAY84437 (1-384)
 QY 95 ATGAACGGAGGAGGAGTTCTTTTACCAATTAATAATGTCGGCTAGGGTCGGCGT 154
 DB 197 MetAspProProThrPheThrPheAsnPheAsnGluProTrpValArgGlyArgHis 216
 QY 155 GAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTCACCTGGAC 214
 DB 217 GluThrTyrLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236
 QY 215 -----TTTGGTTATCTTCGAATAAG-----AACGGCTGC 244
 DB 237 GlnArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256
 QY 245 CACGTGGAATGTCTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGC 304
 DB 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspLeuAspGlnAsp 276
 QY 305 TACCGCGTCACCTGGTTTCACTCTCGAGCCCTCTACGACTGTGCCGACATCTGCC 364
 DB 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296
 QY 365 GACTTTCTGCGAGGAGCAACCCCACTCAGTCTGAGGATCTTCAACGGCGCGCTCTACTTC 424
 DB 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTrp--- 315
 QY 425 TGTGAGGACCGCAAGCTGAGCCCGGCTGGCGGCTGACCCGCGGGGTGCA 484
 DB 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334
 QY 485 ATAGCCATCATGACCTTCAAAGATTATTTTACTGTGGAATACTTTTGTAGAAACCAT 544
 DB 335 IleSerIleLeuThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354
 QY 545 GAAAGAACTTTCAAAGCCTGGAAGGCTGCATCAAAATTCAGTTCCTCTCCAGACAG 604
 DB 355 GlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGlnAlaLeuSerGlyArg 374
 QY 605 CTTGCGGCGATCCTT 619
 DB 375 LeuArgGlyIleLeu 379
 RESULT 11
 AAU31902
 ID AAU31902 standard; Protein; 107 AA.
 XX
 AC AAU31902;
 XX
 DT 18-DEC-2001 (first entry)

CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC The invention on their biological activities, polypeptides and nucleotides of
CC conditions, e.g., by protein or gene therapy, treating or ameliorating medical
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 134 AA;

Alignment Scores:
Pred. No.: 3,65e-34 Length: 134
Score: 383.50 Matches: 88
Percent Similarity: 71.64% Conservative: 8
Best Local Similarity: 65.67% Mismatches: 31
Query Match: 7.77% Indels: 7
DB: 22 Gaps: 2

US-09-966-880A-7 (1-2818) x ABB12093 (1-134)

QY 2117 TTTCTGAGACAGATCTTCTGCTTTCGCCCGGCTGGAGTGAACGGCAGCATCTCAGC 2058
Db 1 PheLeuArgProSerPheAlaLeuValProGlnAlaGlyValGlnTrpCysAlaLeuSer 20
QY 2057 TTACTGCACTCCACCTCCTCGGTTTCAAGCGATTCTCTGCTCAGCTCCAGTAGC 1998
Db 21 TrpLeuGlnProSerProArgPheLys***PheSerCysLeuSerLeuProSerSer 40
QY 1997 TGGGATTACAGGTGCTGTACACGCGCTGGCTAAT---TTTGCATTTTGGTACAGAC 1941
Db 41 TrpAspTyrArgHisValProArgProAlaAsnPhePheValLeuLeu-ValGluTh 60
QY 1940 GGGGTTTCCCATGTGGCCAGGCTGTCTCAAACTCTGTACACAGGTGATCGCCCGG 1881
Db 60 rGlyPheLeuHisValGlyGlnAlaGlyHisGluProLeuThrSerGlyAspProAl 80
QY 1880 CTCGCCCTCCCAAGTGTGGGATTACAGCGGTGAGCCACGCCCGCTCTCTCTT 1821
Db 80 aSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisGlnAlaTrpProSer----- 98
QY 1820 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1761
Db 99 -----Phe-IlePheSerArgAspThrValLeuLeuLeuCysSerGlyTrpSera 116
QY 1760 GAACACQCTGTGTTCAGACCATCTCTCTCTCTCAACCTC 1723
Db 116 rgThrSerGlyLeuGlnSerAlaCysLeuSerLeu 128

RESULT 13
AAB90650
ID AAB90650 standard; Protein: 239 AA.
XX
AC AAB90650;

XX
DT 01-JUN-2001 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 193.

KW Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW neotropic; anticonvulsant; antialzheimers; antiparkinsonian;
KW antimicrobial; vulnery; vaccine; gene therapy; cancer;
KW protein coordinate data; infection.
XX Homo sapiens.
OS
XX WO200121658-A1
XX 29-MAR-2001.
XX
XX 22-SEP-2000; 2000WO-US26013.
XX
XX 24-SEP-1999; 99US-0155709.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
XX Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
XX Young PE, Wei P, Florence KA;
XX
XX WPI; 2001-235311/24.
XX
XX Nucleic acids encoding 32 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX Disclosure; Page 858; 890pp; English.

CC The present sequence is provided in a specification relating to nucleic
CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic
CC acid molecules and polypeptides may be used in the prevention, diagnosis
CC and treatment of diseases such as immune disorders (e.g. multiple
CC sclerosis, systemic lupus erythematosus and human immuno-deficiency virus
CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
CC Gaucher's disease), cardiovascular diseases (e.g. cancers and
CC Chaga's cardiomyopathy and coronary arteriosclerosis), angioenic
CC disorders (e.g. corneal graft neovascularisation and diabetic
CC retinopathy), neurological disorders (e.g. Huntington's chorea,
CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic
CC acid molecules may be used to produce the secreted polypeptides. The nucleic
CC also be used as DNA probes in diagnostic assays to detect and quantitate
CC the presence of similar nucleic acid sequences in samples. The
CC polypeptides may be used as antigens in the production of antibodies and
CC in assays to identify modulators of their expression and activity.

SQ Sequence 239 AA;

Alignment Scores:
Pred. No.: 8,63e-34 Length: 239
Score: 381.50 Matches: 87
Percent Similarity: 72.39% Conservative: 10
Best Local Similarity: 64.93% Mismatches: 27
Query Match: 7.73% Indels: 10
DB: 22 Gaps: 3

US-09-966-880A-7 (1-2818) x AAB90650 (1-239)

QY 2189 CCTTGAATGCTTCCCA-----TCCTTCTCTCCCAATATTGTCCTT 2145
Db 87 ProCysCysSerSerAlaTrpProGluGlySerPheArgProPheGlnMeCASHLeu 106
QY 2144 TCTCTCTCTCTCTCTTT-----TTTTTTTTTTCTGAGACAGAGTCTTTCCTT 2094
Db 107 PheSerPheLeuSerPhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 11
QY 2093 GTCGCCAGGCTGGAGTCAACGGCAGCATCTACGTTACTGCAACCTCCACCTCTCGG 2034
Db 126 UserProArgLeuGluCysSerSerAlaIleSerAlaHisCysAsnLeuArgLeuProG 146

QY 2033 TTCAAGCGATTCTCTGCTCAGCCCTCCCAAGTAGCTGGATTACAGGTGCCTGCTACCA 1974
Db 146 ySerAsnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisI 166
QY 1973 CGCCTGGCTAATTTTGCATTGTGAGTACAGACGGGTTTGGCATGTTGGCCAGCTGG 1914
Db 166 sAlaArgGlnIlePheValPheLeuValGluThrGlyPheCysHisValGlyGlnAlaG 186
QY 1913 TCTCAACCTCTGACCAAGGTGATCCCGCGCTCGGCTCCCAAGTGCCTGGGATTAC 1854
Db 186 yLeuGluLeuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIleI 206
QY 1853 AGCGTGAGCCACCGCCGCGCC---TCTCTCTTTCTT 1817
Db 206 eGlyValSerHisArgAlaArgProGlySerValPheLeu 219
RESULT 14
AAB90647
ID AAB90647 standard; Protein; 310 AA.
AC AAB90647;
XX
DT 01-JUN-2001 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 190.
XX
KW Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW neotropic; anticonvulsant; antihelmets; antiparkinsonian;
KW antimicrobial; vulnerary; vaccine; gene therapy; cancer;
KW protein coordinate data; infection.
XX
OS Homo sapiens.
XX
PN WO200121658-A1.
XX
PD 29-MAR-2001.
XX
PF 22-SEP-2000; 2000WO-US26013.
XX
PR 24-SEP-1999; 99US-0155709.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J., Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI Young PE, Wei P, Florence KA;
XX
XX WPI; 2001-235311/24.
XX
DR Nucleic acids encoding 32 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Disclosure; Page 856; 890pp; English.
XX
CC The present sequence is provided in a specification relating to nucleic
CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic
CC acid molecules and polypeptides may be used in the prevention, diagnosis
CC and treatment of diseases such as immune disorders (e.g. multiple
CC sclerosis, systemic lupus erythematosus and human immuno-deficiency virus
CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiolegnic
CC disorders (e.g. corneal graft neovascularisation and diabetic
CC retinopathy), neurological disorders (e.g. Huntington's chorea,
CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic
CC acid molecules may be used to produce the secreted polypeptides. They may
CC also be used as DNA probes in diagnostic assays to detect and quantitate
CC the presence of similar nucleic acid sequences in samples. The
CC polypeptides may be used as antigens in the production of antibodies and

CC in assays to identify modulators of their expression and activity.
XX
SQ Sequence 310 AA;
Alignment Scores: 1e-33 Length: 310
Pred. No.: 381.50 Matches: 87
Score: 72.39% Conservative: 10
Percent Similarity: 64.93% Mismatches: 27
Best Local Similarity: 7.73% Indels: 10
Query Match: 22 Gaps: 3
DB:
US-09-966-880A-7 (1-2818) x AAB90647 (1-310)
QY 2189 CTTTGAATGCTTCCCA-----TCCTTCTCTCCAAATATTGTCTCTT 2145
Db 158 ProCysCysSerSerAlaTrpProGluGlySerPheArgProPheGlnMetAsnLeu 177
QY 2144 TCTCTCTCTCTCTCTTTT-----TTTTTTTTTCTGAGACAGAGTCTTGTCTT 2094
Db 178 PheSerPheLeuSerPhePhePhePheLeuArgTrpSerLeuThr-Le 197
QY 2093 GTCGCCAGCTGGAGTGCAACGCGAGTCTCAGCTTACTGCAACCTCCACCTCCTGGG 2034
Db 197 uSerProArgLeuGluCysSerSerAlaIleSerAlaHisCysAsnLeuArgLeuProGl 217
QY 2033 TTCAAGCGATTCTCTGCTCAGCTCCCAAGTAGCTGGATTACAGGTGCCTGCTACCA 1974
Db 217 ySerAsnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisI 237
QY 1973 CGCCTGGCTAATTTTGCATTGTGAGTACAGACGGGTTTGGCATGTTGGCCAGCTGG 1914
Db 237 sAlaArgGlnIlePheValPheLeuValGluThrGlyPheCysHisValGlyGlnAlaG 257
QY 1913 TCTCAACCTCTGACCAAGGTGATCCCGCGCTCGGCTCCCAAGTGCCTGGGATTAC 1854
Db 257 yLeuGluLeuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIleI 277
QY 1853 AGCGTGAGCCACCGCCGCGCC---TCTCTCTTTCTT 1817
Db 277 eGlyValSerHisArgAlaArgProGlySerValPheLeu 290
RESULT 15
AAB42917
ID AAB42917 standard; Protein; 165 AA.
XX
AC AAB42917;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF3681 polypeptide sequence SEQ ID NO:5362.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; neotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
XX 05-OCT-2000.
XX

Search completed: June 14, 2003, 18:25:26
Job time : 125.983 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	370	7.5	673	4	F40201		artifact-warning s
C 2	348	7.1	613	4	C40201		artifact-warning s
C 3	340	6.9	627	4	A40201		artifact-warning s
C 4	339	6.9	613	4	C40201		artifact-warning s
C 5	330	6.7	673	4	F40201		artifact-warning s
C 6	329.5	6.7	627	4	A40201		artifact-warning s
C 7	280.5	5.7	301	4	B40201		artifact-warning s
C 8	270	5.5	597	4	E40201		artifact-warning s
C 9	265.5	5.4	597	4	E40201		artifact-warning s
C 10	254	5.1	196	2	I38022		hypothetical prote
C 11	246.5	5.0	301	4	G40201		artifact-warning s
C 12	245	5.0	116	2	B01233		phorbolin I - huma
C 13	243	4.9	100	2	A46010		X-linked retinopat
C 14	225	4.5	236	2	A53853		apolipoprotein B m

Alignment Scores:		
Pred. No.:	2,576-24	Length:
Score:	339.00	Matches:
Percent Similarity:	51.32%	Conservative:
Best Local Similarity:	43.92%	Mismatches:
Query Match:	6.85%	Indels:
DB:	4	Gaps:
		2
		613

US-09-966-880A-7 (1-2818) x C40201 (1-613)

	Qy	1741	TGCGTTCAACAC	-----	AGGTGTTCAAGCCGACCTGGCCACATCAACAGATCCCTGTC	1794
Qy						
Db		121	TrpValAspHisGluAlaArgSerSerArgGrolatrpProThrTrpGlnAsnProIle	140		
Qy		1795	TCTCAAAA	-----		1801

Db		141	Ser-ThrLysAsnThrLysAsn***LeuGlyMetValArgAlaProValValProArgTh	160		
Qy		1801	-----	-----		1801

```

RESULT 5
F40201
artfact-warning sequence (translated ALU class F) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: F40201
R:Claverie, J.M.

```

personal communication, 1992
A:Reference number: A40201
A:Accession: F40201
A:Molecule type: DNA
A:Residues: 1-673 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other p
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading fr
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a port

Alignment Scores:

Pred. No.: 2e-23 Length: 673
Score: 330.00 Matches: 98
Percent Similarity: 52.85% Conservative: 32
Best Local Similarity: 39.84% Mismatches: 59
Query Match: 6.67% Indels: 58
DB: 4 Gaps: 6

US-09-966-880A-7 (1-2818) x F40201 (1-673)

QY	1715	CTGCTGGAGGCTT	GAGGAGGAGGATGCTTGAACACAGGTTCAAGGC-----	1765
DB	57	LeuHisGlyArgGlnArgGlnGluAsnArgLeuAsnProGlyGly***GlyCysSerGlu	76	
QY	1766	-----	-----CAGCCTGGGCAACATAACAAGATCCTGCTCTCAAAAAA	1804
DB	77	ProLysLeuAlaThrAlaLeuProGlyCysGlnSerLysGlyLeuSerGlnLysGln	96	
QY	1805	AAAAAAAAAAAAA	AGACAGAGAGG-----GCCGGGCGT	1839
DB	97	LysGlnSerLysLysLysLysLysLysThrProLysAsnLys*****AlaGlyCys	116	
QY	1840	GTGGCTCAGCGCTGTAATCCAGCACTTTGGAGCGCGAGCGCGGATCACCTGTGG	1899	
DB	117	GlyGlySerArgLeuSerGlnHisPheGlyArgPro-GlyGlyGlnIleThr***G	136	
QY	1900	TCAGGAGTTTGACACAGCCTGGCCAAATGCAAAACCCCTGCTGTACTCAAAATGCAA	1959	
DB	136	yGlnGluPheGluThrSerLeuIleAsnMetValLysLeuCysLeuTyr**LysTyr	156	
QY	1960	AAATTAGCAGCGCTGTAGCAGCGACCTGTAATCCAGCTACTTGGGAGGCTGAGCGAG	2019	
DB	156	easn***ProGlyMetAlaAlaHisAlaCysAsnProSerTyrThrGlyAspArgGlyAr	176	
QY	2020	GAGAATCGCTTGAACCCAGGAGGTGGAGTTCAGTAACTGATCGTGGCGTTGCACT	2079	
DB	176	gargileAla**ThrGlnGluValGluAlaValSer-GlnAsnLeuProLeuHis	196	
QY	2080	CGAGCTGGGACAGACAGACTCTCTCTCAGAAAAA*****AAGAGAGAGAG	2139	
DB	196	erSerArgGly-ValArgAlaArgAlaTyrLeuLysAsnLysAsnLysAlaLysLys	215	
QY	2140	ACAGAAGA-----	2148	
DB	216	LysLysLysProProLysThrLys*****LeuGlyValValAlaHisAlaCysHis	235	
QY	2149	GAACAATATTGGAGAGAGGATGGGAAGCATTTGCAAGGAAATGTGCTTTATCCAAAC	2208	
DB	236	ProSerThrLeuGlyAspGlnGluGlyArgSerLeuGluValArg-----	250	
QY	2209	AAATGTAAAGGAGCCAAATAAGGATCCCTATTGTCCTTTTGGTGCTCTATTGTCCTTA	2268	
DB	251	-----SerLeuArgProAla**SerThrTrp***AsnCys	262	
QY	2269	ACAACCTGCTTTCACAGTCAGAAAAAT-----ATTCAGAAATACCATATCCTGTGCG	2322	
DB	263	ValSerIle-----LysAsnThr***IleSerGlnGluTrpArgProMetPro	278	
QY	2323	TTATTACCTPAGC	2334	
DB	279	ValIleProAla	282	

RESULT 6

A40201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA

A:Residues: 1-627 <CLA>

R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentl
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: this "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:
Pred. No.: 2.2e-23 Length: 627
Score: 329.50 Matches: 90
Percent Similarity: 67.52% Conservative: 16
Best Local Similarity: 57.33% Mismatches: 35
Query Match: 6.66% Indels: 17
DB: 4 Gaps: 3

US-09-966-880A-7 (1-2818) x A40201 (1-627)

QY	1715	CTGCTGGAGGCTT	GAGGAGGAGGATGCTTGAACACAGGTTGT-----	1758
DB	56	LeuLeuGlyArgLeuArgGlnGluAsnGlyValAsnProGlyGlyGlyAlaCysSerGlu	75	
QY	1759	-----TCAAGG-----	-----CCAGCCTGGGCAACATAACAAGATCCTCTCTCAAAAA	1803
DB	76	ProArgSerArgHisCysThrProAlaLeuAlaThrGluArgAspSerValSerGluLys	95	
QY	1804	AAAAAAAAAAAAA	AGAAAGA---CAGAGGCGCGCGGCTGGTGGCTCAGCGCTGTAAATCC	1860
DB	96	AsnLysAsnLysLys***Lys*****AlaGlyArgGly--LeuThrProValIleP	115	
QY	1861	CAGCACTTTGGAGGCGCGGATCACCTGTGCTCAGGAGTTTGACACAGCGCT	1920	
DB	115	roAlaLeuTrpGluAlaLysAlaGlyGlySerArg--GlyGlnGluIleGluThrIleLe	134	
QY	1921	GGCCAACATGGCAAAACCCCGTCTGCTACTCAAAATGCAAAATTTAGCCAGGCGTGTAGC	1980	
DB	134	uAlaThrThrValLysProArgLeuTyr***LysTyrLysLysLeuAlaGlyArgSerAl	154	
QY	1981	AGGCACCTCTAATCCAGCTACTTGGAGGCTGAGCGAGGAGGAGGAGGAGGAGGAGGAG	2040	
DB	154	aglyAlaCysSerProSerTyrLeuGlyGly***GlyArgArgMetAla**ThrArgG	174	
QY	2041	GGTGGAGGTTCAGTAAAGCTGAGATCGTCCCTTTCAGCTCCAGCTGGCGCAACAGCA	2100	
DB	174	uAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuGlnProTrpArgGln-SerG	194	
QY	2101	AGACTCTGCTCAGAAAAA*****AAGAGAGAGAGAGAG	2141	
DB	194	luThrProSerArgLysLysIleLysThrLysAsnLysLys	207	

RESULT 7

B40201
artifact-warning sequence (translated ALU class B) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: B40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: B40201
A:Molecule type: DNA
A:Residues: 1-301 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentl
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-116 <NAD>

A:Cross-references: EMBL:U03891; NID:9436940; PIDN:AAA03706.1; PID:9436941

C:Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1

Alignment Scores:
Pred. No.: 2,98e-15 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 4.95% Indels: 10
DB: 2 Gaps: 3

US-09-966-880A-7 (1-2818) x G01233 (1-116)

```
QY 305 TACCGCGTCACTGGTTACCTCCCTGGAGCCCTGTCAGCAC-----TGTGCCCGACAT 358
D 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26
QY 359 GTGGCGGACTTCTGCGAGGAGCCCAACCTCAGCTCAGGATCTTCCACCGCGCGCTC 418
D 27 ValArgAlaPheLeuGlnGlnThrHisValArgLeuProIlePheAlaAlaArgIle 46
QY 419 TACTTCTGTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTGGCGGCTGCAC 469
D 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
QY 470 CGCGCGGGGTGCAATATGCCATCATGACCTTCAAGATATTTTACGTGGGATACT 529
D 62 AspAlaGlyAlaGlnValSerIleMetThrTrpAspGluPheGluTyrCysTrpAspThr 81
QY 530 TTGTGTAGAAACCATGAAAGAACTTTCAAGGCTGGGAAGGCTGCATGAAATTCAGTT 589
D 82 PheValTyrArgGlnCysProPheGlnProTrpAspGlyLeuGluHisSerGln 101
QY 590 CGTCTCCAGACAGCTTCGGCGCATTCCTT 619
D 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111
```

RESULT 13

A46010

X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A46010

R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.

Genomics 15, 467-471, 1993

A:Title: Identification and partial characterization of a candidate gene for X-linked re

A:Reference number: A46010; MUID:93224131; PMID:8468040

A:Accession: A46010

A:Molecule type: nucleic acid

A:Residues: 1-100 <WON>

A:Cross-references: GB:S58722; NID:9299470; PIDN:AAB26149.1; PID:9299471

A>Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBI:129340)

Alignment Scores:
Pred. No.: 4,55e-15 Length: 100
Score: 243.00 Matches: 63
Percent Similarity: 65.18% Conservative: 10
Best Local Similarity: 56.25% Mismatches: 27
Query Match: 4.92% Indels: 12
DB: 2 Gaps: 3

US-09-966-880A-7 (1-2818) x A46010 (1-100)

QY 2130 TTTTCTTTTCTTGAGACAGAGTCTTGCTTGTGCGCCAGGCTGGAGTGAACG 2071

D 1 PhePhePhePhePheGluThrGluSerCysSer-ValAlaGluAlaGlyValGlnTr 20

QY 2070 GCACGATCTCACTTACTGCAACCTCCACCTCCTGGGTTCAACGAGTTCCTCGCTCAG 2011

D 2070 GCACGATCTCACTTACTGCAACCTCCACCTCCTGGGTTCAACGAGTTCCTCGCTCAG 2011

```
D 20 pCysAspLeuGlySerLeuLysSerProProGlySerSerAspSerProAlaSerA 40
QY 2010 CCTCCCAAGTAGCTGGATTACAGGTCCCTGCTACACACGCTGGCTAATTTTTCATTTT 1951
D 40 lAserArgValAlaGlyIleThrGlyMetHisHisThrGlnLeuIlePheValPhe 60
QY 1950 GAGTACAGACGGGTTTTCATGTTGGCAGCTGGTCTCAAACTCCTCACCACACGGTG 1891
D 60 euValGluThrGly---SerHisMetGlnLeuSerAspSerThrLeuValIleThr---- 77
QY 1890 ATCCGCCGCGCTGGCCTCCCAAGTGTGGGATTACAGCGGTGAGCCACCGCCCGCGC 1831
D 78 -----ThrAlaGlnAsnAlaLysIleThrAla-----ArgAlaProA 90
QY 1830 CCTCTCTCTTCTTTTCTTTTCTTTTCTTTT 1799
D 90 rgAspLeuPhePhePhePhePhePhePhe 100
RESULT 14
A53853
apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 18-Aug-2000
C:Accession: A53853
R:Yamanaka, S.; Poksay, K.S.; Balestra, M.E.; Zeng, G.Q.; Innerarity, T.L.
J. Biol. Chem. 269, 21725-21734, 1994
A:Title: Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc mod
ributed.
A:Reference number: A53853; MUID:94342367; PMID:8063816
A:Accession: A53853
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-236 <YAM>
A:Cross-references: GB:U10695; NID:g506180; PIDN:AAA56718.1; PID:g506181
C:Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1
C:Keywords: hydrolase; zinc
Alignment Scores:  
Pred. No.: 3,12e-13 Length: 236  
Score: 225.00 Matches: 49  
Percent Similarity: 57.25% Conservative: 26  
Best Local Similarity: 37.40% Mismatches: 48  
Query Match: 4.55% Indels: 8  
DB: 2 Gaps: 4
US-09-966-880A-7 (1-2818) x A53853 (1-236)
QY 149 CGCGGTGAGACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTCA 208
D 33 ArgLysGluAlaCysLeuLeuTyrGluIleLysTrpGlyAlaSerSerLysThrTrpArg 52
QY 209 CTGGACTTTGGTTATCTTCGCAATAGAACGGCTGC---CACGTGAANTTGTCTTCCTC 265
D 53 SerSer-----GlyLysAsnThrThrAsnHisValGluValAsnPheLeu 67
QY 266 ---CGCTACATCTCGGACTGGGACCTAGACCTCGGCTGGCTACCGCGTCACTGGTTC 322
D 68 GluLysLeuThrSerGluGlyArgLeuGlyProSerThrCysCysSerIleThrTrpPhe 87
QY 323 ACCTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACATTTCTGCGAGGAAC 382
D 88 LeuSerTrpSerProCysTrpGluCysSerMetAlaIleArgGluPheLeuSerGlnHis 107
QY 383 CCCAACCTCAGTCTGAGGATCTTCCCGCGGCTCTACTTCTGTGAGGACCCCAAGGCT 442
D 108 ProGlyValThrLeuIleIlePheValAlaArgLeuPheGlnHisMetAspArg--- 126
QY 443 GAGCCCGAGGGCTGGCGGGTGCACCGCGCGGGTGCACCAATAGCCATCATGACCTTC 502
D 127 AsnArgGlnGlyLeuLysAspLeuValThrSerGlyValThrValArgValMetSerVal 146
QY 503 AAAGATTATTTTACTGCTGGAATACTTTTGTGA 535
D 503 AAAGATTATTTTACTGCTGGAATACTTTTGTGA 535
```

```

Db      147 SerGluTyrCysTyrCysTrpGluAsnPheVal 157
RESULT 15
S41044
chromosomal protein - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S41044
R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A:Title: A new chromosomal protein essential for mitotic spindle assembly.
A:Reference number: S41044; MUID:94166884; PMID:8121495
A:Accession: S41044
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-418 <YEO>
A:Cross-references: GB:L26953; NID:g537529; PIDN:AAB68050.1; PID:g537530
C:Superfamily: human 48.2K chromosomal protein
C:Keywords: Chromosomal protein

Alignment Scores:
Pred. No.:      4.9e-13      Length:      418
Score:          223.50      Matches:      64
Percent Similarity: 51.61%      Conservative: 32
Best Local Similarity: 34.41%      Mismatches: 36
Query Match:      4.52%      Indels:      54
DB:              2          Gaps:      7

US-09-966-880A-7 (1-2818) x S41044 (1-418)

QY      1794 CTCTCAAAAAAAAAAAAAA----- 1814
Db      127 IleSerGlnLysLysLysArgGlyIleTyrHisLysAsnAsnIleArgIleIleLeuPhe 146
QY      1815 AAAAGAAGAGAGAGCGCGCGGTGGCTGACGCTGTAATCCCA----- 1862
Db      147 LeuProGlnAlaHisGlyArgAspPheTyrVal---ProIleLeuProPheThrGlnSer 165
QY      1863 -----GCACCTTTGGAGCGCGCGCGGATCACCTGTGTC 1901
Db      166 TyrValAspTTPGlyArgTrpLeuIleTrpGluAlaLysAlaGlyGluSerLeuGluVal 185
QY      1902 AGGAGTTTGACACAGCGCTGCCCAACATGCCAAAC----- 1937
Db      186 ArgSerSerArgProAlaSerGlnSerArgArgAsnSerValSerThrLysAsnIleLys 205
QY      1938 -----CCCGTCTGTACTCAAAATGCAAAATAGCCAGCGCTGGTAGCAGCACCT--- 1988
Db      206 IleSerProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyGly 225
QY      1989 -----GTAATCCAGCTACTGGAGGCTGAGCGAGAGAAATCGCTTGAACCCAGG 2039
Db      226 ValHisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAsp-ProG1 245
QY      2040 AGGTGGAGTTGAGTAAGCTGAGATCGTCCCTTGCCTCAGCTCCAGCTGGCGCAAGAGC 2099
Db      245 YGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThr--Ser 264
QY      2100 AAGACTCTGTCTCAGAAAAA----- 2147
Db      265 GluThrValSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 284
QY      2148 -----AGACATATTTGGGAGAGAGGATGGGAAGCATTCGAAGGAAATGTGC 2198
Db      285 IleAsnAlaSerThrLeuPhe-----HisValLeuThrArgIlePheCys 299
QY      2199 TTTATCCAAACA 2210
Db      300 TyrLysGlnLys 303

```

Search completed: June 14, 2003, 18:56:28
Job time : 98.1094 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:11:13 ; Search time 35.3473 Seconds
(without alignments)
6613.252 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 4948
Sequence: 1 agagaacatcattaatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/sgn2.1/USPTO-spool/US09966880/runat_14062003_175524_10304/app.query.fasta_1.9493
-DB=SwissProt_40 -GFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 -ECGN_1_1_196.@runat_14062003_175524_10304 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Match	Length	ID	Description
C 1	498	10.1	593	1 ALU7_HUMAN	P39194 homo sapien
C 2	484	9.8	593	1 ALU6_HUMAN	P39193 homo sapien
C 3	457	9.2	593	1 ALU7_HUMAN	P39194 homo sapien
C 4	430	8.7	593	1 ALU6_HUMAN	P39193 homo sapien
C 5	420.5	8.5	591	1 ALU1_HUMAN	P39188 homo sapien
C 6	416.5	8.4	591	1 ALU1_HUMAN	P39188 homo sapien
C 7	414	8.4	591	1 ALU8_HUMAN	P39195 homo sapien
C 8	382.5	7.7	591	1 ALU8_HUMAN	P39195 homo sapien
C 9	372	7.5	585	1 ALU5_HUMAN	P39192 homo sapien
C 10	368.5	7.5	587	1 ALU2_HUMAN	P39189 homo sapien
C 11	366	7.4	585	1 ALU5_HUMAN	P39192 homo sapien
C 12	364	7.4	382	1 PHB3_HUMAN	Q9uh17 homo sapien
C 13	363.5	7.3	199	1 PHB1_HUMAN	P39141 homo sapien
C 14	359.5	7.3	587	1 ALU3_HUMAN	P39190 homo sapien
C 15	358	7.2	587	1 ALU3_HUMAN	P39190 homo sapien
C 16	356	7.2	587	1 ALU2_HUMAN	P39189 homo sapien
C 17	339	6.9	190	1 PHB2_HUMAN	Q9ue74 homo sapien
C 18	295	6.0	603	1 ALU4_HUMAN	P39191 homo sapien

19	288.5	5.8	603	1	ALU4_HUMAN	P39191 homo sapien
20	225	4.5	236	1	ABME_RABIT	P47855 oryctolagus
21	223.5	4.5	418	1	YY1_HUMAN	P49646 homo sapien
22	221	4.5	236	1	ABME_HUMAN	P41238 homo sapien
23	213.5	4.3	229	1	ABME_MOUSE	P51908 mus musculus
24	199.5	4.0	229	1	ABME_RAT	P38483 rattus norv
C 25	182.5	3.7	841	1	NEK4_HUMAN	P51957 homo sapien
C 26	143.5	2.9	124	1	YY3_HUMAN	P20931 homo sapien
27	134	2.7	447	1	KBF3_HUMAN	Q04860 homo sapien
C 28	130	2.6	369	1	K2R2_HUMAN	P21731 homo sapien
29	128.5	2.6	881	1	PRP2_HUMAN	Q99599 homo sapien
C 30	120	2.4	619	1	REL_HUMAN	Q04864 homo sapien
C 31	119	2.4	629	1	Z195_HUMAN	Q14628 homo sapien
C 32	108	2.2	1191	1	ZN91_HUMAN	Q05481 homo sapien
C 33	106	2.1	1371	1	UBP1_HUMAN	Q04966 homo sapien
C 34	98.5	2.0	369	1	CT19_YEAST	Q02732 saccharomyc
C 35	94	1.9	504	1	AT1N_HSVBP	P30020 bovine herp
36	92	1.9	2476	1	ATRX_MOUSE	Q61687 mus musculus
37	91	1.8	1402	1	IF4G_RABIT	P41110 oryctolagus
38	90.5	1.8	522	1	NU62_HUMAN	P37198 homo sapien
39	89.5	1.8	976	1	BUL1_YEAST	P48524 saccharomyc
40	89	1.8	354	1	WN14_CHICK	O42280 gallus gall
C 41	89	1.8	634	1	YCK3_EUGER	P31916 euglena gra
C 42	89	1.8	1064	1	ISK5_HUMAN	Q9nq38 homo sapien
C 43	88.5	1.8	619	1	VALL_YEAST	P38085 saccharomyc
C 44	88	1.8	735	1	PSAB_CHLRE	P09144 chlamydomon
45	88	1.8	914	1	GNDS_HUMAN	Q12967 homo sapien

ALIGNMENTS

RESULT 1					
ID	ALU7_HUMAN	STANDARD	PRT	593 AA.	
AC	P39194;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	16-Oct-2001 (Rel. 40, Last annotation update)				
DE	Alu subfamily SQ sequence contamination warning entry.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=95021758; PubMed=7935834;				
RA	Claverie J.-M., Makalowski W.;				
RT	"Alu alert.";				
RL	Nature 371:752-752(1994).				
RN	[2]				
RP	CONCEPT.				
RP	MEDLINE=92241891; PubMed=1572661;				
RA	Claverie J.-M.;				
RT	"Identifying coding exons by similarity search: alu-derived and other				
RT	potentially misleading protein sequences.";				
RL	Genomics 12:838-841(1992).				
RN	[3]				
RP	ALU FAMILIES CLASSIFICATION.				
RP	MEDLINE=88333009; PubMed=3138422;				
RA	Quentin Y.;				
RT	"The Alu family developed through successive waves of fixation				
RT	closely connected with primate lineage history.";				
RL	J. Mol. Evol. 27:194-202(1988).				
RN	[4]				
RP	ALU FAMILIES CLASSIFICATION.				
RP	MEDLINE=91178815; PubMed=1706781;				
RA	Jurka J., Milosavljevic A.;				
RT	"Reconstruction and analysis of human Alu genes.";				
RL	J. Mol. Evol. 32:105-121(1991).				
CC	-!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE				
CC	THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING				
CC	CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX				
CC	FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC	EMBL; U14572; -; NOT_ANNOTATED_CDS.
DR	Hypochemical protein.
FW	DOMAIN 1 97
FT	DOMAIN 101 196
FT	DOMAIN 200 295
FT	DOMAIN 299 395
FT	DOMAIN 399 494
FT	DOMAIN 498 593
SQ	SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:		
Pred. No.:	2 6e-30	Length: 593
Score:	430.00	Matches: 95
Percent Similarity:	55.73%	Conservative: 12
Best Local Similarity:	49.48%	Mismatches: 35
Query Match:	8.69%	Indels: 50
DB:	1	Gaps: 1
US-09-966-880A-7 (1-2818) x ALU6_HUMAN (1-593)		

RESULT 5

ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)

FT	DOMAIN	100	195	FRAME-2.
FT	DOMAIN	199	294	FRAME-3.
FT	DOMAIN	298	393	FRAME-4.
FT	DOMAIN	397	492	FRAME-5.
FT	DOMAIN	496	591	FRAME-6.
SQ	SEQUENCE	591 AA;	63790 MW;	665D395735519D95 CRC64;

Alignment Scores:			
Pred. No.:	1.84e-29	Length:	591
Score:	420.50	Matches:	116
Percent Similarity:	43.88%	Conservative:	13
Best Local Similarity:	39.46%	Mismatches:	81
Query Match:	8.52%	Indels:	84
BD:	1	Gaps:	5

US-09-966-880A-7 (1-2818) x ALUL_HUMAN (1-591)			
Qy	2359	TCYGCATCTTCATTCGAAAGGTGCTAGGTAATAACGCCACAGGATATGGTTATTCT	2300
Db	307	SerProArgLeuGluCysSerGly-----AlaIleThrAla-----	318
Qy	2299	GAATATTTTCTCAGCTCAAAAGACAGTGTGTAGGCACAAT-	2258
Db	319	-----HisCysSerLeuAspLeuProGlySerSerAspProProAlaSerAla	334
Qy	2257	-----AGACACCAAAAGACAAATAGGATGCCCTTAT	2225
Db	335	SerArgValAlaGlyThrThrGlyAlaArgHisAlaArgLeuIlePheValPhePhe	354
Qy	2224	TGGCTCCTTACATTTTGTGGTAAGCAACAATTCCTTGCAGTCTCCCATCCTTCT	2165
Db	355	ValGluThrGlyPheHisIstyrValAlaGlnAlaGlyLeuGluLeuGlySerSerAsp	374
Qy	2164	CTCCCA-----AATATTCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2126
Db	375	ProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg	394
Qy	2125	TTTTTTTTTTCTGAGACAGTCTTCTCTGTGCGCCAGGCTGGAGTGCAAGCGCAGC	2066
Db	395	*****PhePhe***AspArgValSerLeu-CysArgProGlyTrpSerAlaValAla	414
Qy	2065	ATCTCAGCTTACTGCAACCTCCACTCTCTGGGTCAAGCGATCTCCTCGCTCAGCTCC	2006
Db	414	gSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuProGlnPro	434
Qy	2005	CAACTAGCTGGATTACAGTGGCTGCTACCGCTGGCTAATTTTGCATTTTGAGTA	1946
Db	434	oglu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu	454
Qy	1945	CAGACGGGTTTGGCATTTGGCCAGGTGTCTCAAACTCCTGACCACAGGTGATCCG	1886
Db	454	*ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGlnValIle	474
Qy	1885	CCGGCTGCGCTCCCAAAGTGTGGATTACAGCGCTGAGCCACCCCGCCGCCCTCT	1826
Db	474	uProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly****	494
Qy	1825	CTCTTCTCTT-----	1816
Db	494	***PhePheGluThrGlySerArgSerValAlaGlnAlaGlyValGlnTrpArgAsp	514
Qy	1816	-----	1816
Db	514	sGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuPro	534
Qy	1815	-----TTTTTTTTTTTTTTTTTTTGAGAGA	1793
Db	534	rSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheCysIlePheCysArg	554
Qy	1792	CAGGATCTGTATTGTCGCCAGGCTGCCCTTGACACCTGTGTTCAAGCCATCTCCCT	1733
Db	554	pglyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSer	574

CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U14574; -; NOT_ANNOTATED_CDS.
CC Hypothetical protein.
CC FT DOMAIN 1 96 FRAME-1.
CC 100 195 FRAME-2.
CC 199 294 FRAME-3.
CC 298 393 FRAME-4.
CC 397 492 FRAME-5.
CC 496 591 FRAME-6.
CC SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Alignment Scores:
Pred. No.: 4,67e-26 Length: 591
Score: 382.50 Matches: 103
Percent Similarity: 68.12% Conservative: 6
Best Local Similarity: 64.38% Mismatches: 32
Query Match: 7.73% Indels: 20
DB: 1 Gaps: 5

US-09-966-880A-7 (1-2818) x ALU8_HUMAN (1-591)

QY 1679 AAGCTGCGAACCGT---GGCACAGCTCATAGTTCTAGCTGCTGGAGGTGAGGAGG 1735
Db 143 LysLeuAlaGlyArgGlyAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGln 162
QY 1736 GAGGATGCTTCACACAGGTGT-----TCAGG----- 1764
Db 163 GluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHisCysThr 182
QY 1765 CCAGCTGGGCAACATAACAGATCTGTCTCTCAAAAAAAGAAAAAAGAAAAA 1824
Db 183 ProAlaThrAlaThrGluArgAspSerValSer-----LysLys*****Arg 199
QY 1825 GAGGGCGGGCGGTGGTGGCTGACCGCTGTAATCCAGCACTTTGGAGGCCGCGG 1884
Db 200 AlaGluAlaGlyArgGlySerArgLeu***SerGlnHisPheGly-----Arg 216
QY 1885 CGGGATCACCTGTGTCAGGAGTTTGACACAGCCCTGGCCACATGGCAAAACCCGCT 1944
Db 217 -LysileThr***GlyGlnGluPheGluThrSerLeuAlaAsnMetValLysProArgLe 236
QY 1945 GTACTCAAAATGCAAAATATTAGCCAGCGTGGTAGCGACCTCTAATCCAGCTACTT 2004
Db 236 uTyR***LysTyRlysAsn***ProGlyValValAlaArgAlaCysAsnProSerTySe 256
QY 2005 GGGAGGCTGAGCAGGAGAGTTCGTTGAACCCAGGAGGTGGAGGTGAGTAAAGCTGAGA 2064
Db 256 rGlyGly***GlyArgArgIleAla***ThrArgGluAlaGluValAlaValSerArgAs 276
QY 2065 TCGTGCCCTGCTCACTCCAGCTGGCGGACAGACAGCACTGTCTCAGAAAAA 2120
Db 276 pArgAlaThrAlaLeuGlnProGlyArgGln-SerGluThrProSerGlnLysLys 294

RESULT 9

ALU5_HUMAN
ID ALU5_HUMAN STANDARD: PRT: 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS; VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS; ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; U14571; -; NOT_ANNOTATED_CDS.

Hypothetical protein.

DOMAIN 1 95

FRAME-1.

```

FT DOMAIN 99 193 FRAME-2.
FT DOMAIN 197 291 FRAME-3.
FT DOMAIN 295 389 FRAME-4.
FT DOMAIN 393 487 FRAME-5.
FT DOMAIN 491 585 FRAME-6.
SQ SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.: 4.06e-25 Length: 585
Score: 372.00 Matches: 96
Percent Similarity: 52.45% Conservative: 11
Best Local Similarity: 47.06% Mismatches: 70
Query Match: 7.54% Indels: 27
DB: 1 Gaps: 4

US-09-966-880A-7 (1-2818) x ALU5_HUMAN (1-585)
QY 2237 AGGGATCCCTTATTGGCTTCATATTTGTTGGATAAGCACAATTTCTCTTGCAGAGCT 2178
|||||
Db 369 ArgAspPro-----ProAlaSerAla 375
QY 2177 TCCCATCTCTCTCCCAATATATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2118
|||
Db 376 SerGlnSerAlaGlyThrGlyValSerHisArgAlaArg*****PhePhe--- 394
|||
QY 2117 TTTCTGACAGAGCTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2058
|||||
Db 395 -----AspGlyValSerLeuCysArgGlnAlaGlyValGlnTrpArgAspLeuGly 412
|||||
QY 2057 TTATGCAACCTCCACCTCTCTGGGTTCAAGGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1998
|||||
Db 413 SerLeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProSer 432
|||||
QY 1997 TGGGATTACAGTCCCTCTACCGCTGGCTGCTTAATTTTGCATTTTGTAGTACAGAGCGG 1938
|||||
Db 433 TrpAspTyrArgAlaProProArgProAlaAsnPheCysIlePheSerArgAspGly 452
QY 1937 GTTTTGGCATTTGGCCAGGCTGCTCTCAAACTCTCGACACAGGTGATCGCCCGCTGAA 1758
|||||
Db 453 ValSerProCysTrpProGlyTrpSerArgSerLeuAsp--LeuValIleArgProPro 472
QY 1877 GGCTCTCCAAAGTCTGGGATTACAGCGGTGAGCCAGCCAGCCGCGCCCTCTCTCTCTCTCTCTCTCTCT 1818
|||||
Db 472 rGProPolysValLeuGlyLeuGlnAla**AlaThrAlaProGly-----***** 490
QY 1817 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1758
|||||
Db 490 **PhePheGluThrGluSerArgSerValAlaArgLeuGlyCysSerGly-AlaIleSer 509
QY 1757 CACCTGTGT-----TCAAGCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1716
|||||
Db 510 AlaHisCysAsnLeuArgLeuProGlySerSerAspSerProAlaSerAlaSerArgVal 529
QY 1715 GCTAGACTATGACGCTGTGCCAGCTTTCGACGCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1656
|||||
Db 530 AlaGlyThrThrGlyAlaArgHisAlaGlnLeuIlePheValPheLeuValGluThr 549
QY 1655 TCTTAT 1650
|||
Db 550 GlyPhe 551

RESULT 10
ALU2_HUMAN
ID ALU2_HUMAN STANDARD; PRT; 587 AA.
AC P39189;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
OS Alu subfamily SB sequence contamination warning entry.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.,
RT "Alu alert.",
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CONAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U14568; -; NOT_ANNOTATED_CDS.
CC Hypothetical protein.
KW DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 194 FRAME-2.
FT DOMAIN 198 292 FRAME-3.
FT DOMAIN 296 391 FRAME-4.
FT DOMAIN 395 489 FRAME-5.
FT DOMAIN 493 587 FRAME-6.
SQ SEQUENCE 587 AA; 63703 MW; 3EAA3E3E3929203 CRC64;

```



```

Db 165 -----GluProGlyArgArgArgLeuGln***AlaGluLeuAlaProLeuHi 180
QY 1783 CAAGATCCTGCTCTCAAAAGAAAAAAGAGAGAGAGCGCGCGTGGT 1842
Db 180 sSerSerLeuAlaThrGluArgAspSerValSerLysLys*****ProGlyAlaVa 200
QY 1843 GCCTCACCCCTGTAATCCACACACTTTGGGAGCGCGGCGGATCACTCTGGTCA 1902
Db 200 lAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgIleThrArgSerAr 220
QY 1903 GGAGTTTGAGACCGCTGCGCAACATCGCGCAAAACCCGCTGTACTCAAAATCAAAA 1962
Db 220 gAsPArg--AspHisProGlyGlnHisGlyGluThrProSerLeuLeuLysIleGlnLys 239
QY 1963 TTAGCGCGCTGGTAGCAGCCTGTAAATCCACGCTACTTGGGAGCGTGGAGCAGGAG 2022
Db 240 LeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlu 259
QY 2023 AATCGCTTGAACCCAGGAGTGGAGTTGCAGTAAGCTGAGATCGTCCGCTGCACATCCA 2082
Db 260 AsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCysThrPro 279
QY 2083 GCCTGGCGGCAAGCAAGCACTCTGCTCAGAAAAAAGAGAGAGAGAGAGAGA 2142
Db 280 AlaTrp-ArgGln-SerGluThrProSerGlnLysLys*****PhePheLeuArgA 299
QY 2143 GAAAGAGAACAATATTGGGAGAGAGGATGGGAGCAAGCATTCGAAGG 2189
Db 299 rgSerLeuAlaLeu-----SerProGlyTrpSerAlaValAlaArg 312

RESULT 12
PHB3_HUMAN
ID PHB3_HUMAN STANDARD; PRT; 382 AA.
AC Q9UHL7; O95618;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phorboln 3 (APOBEC1-like).
GN APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=20057165; PubMed=10591208;
Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
Hunt S.G., Jones M.C., Kershaw J., Kimberley A.M., King A.,
Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
Shinomiya A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,

```

```

RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kuranashi H., Saitta S., Budarf M.L.,
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkison P., Bodetelch A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RN Nature 402:489-495(1999).
RP SEQUENCE OF 148-382 FROM N.A.
RC TISSUE=Keratinocytes;
RA Madsen P.;
RT "Molecular cloning of phorboln 3.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO APOLOPROTEIN B MRNA EDITING PROTEIN.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-148 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL022318; CAB45270.1; .
DR EMBL; U61084; AAD00090.1; .
DR InterPro; IPR002125; dCMP/cytL_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 2.
KW Hydrolase. 382 AA; 45924 MW; DA6EDD23E8856240 CRC64;
SQ SEQUENCE 382 AA; 45924 MW; DA6EDD23E8856240 CRC64;

Alignment Scores:
Pred. No.: 1,93e-24 Length: 382
Score: 364.00 Matches: 83
Percent Similarity: 57.07% Conservative: 26
Best Local Similarity: 43.46% Mismatches: 62
Query Match: 7.36% Indels: 20
DB: 1 Gaps: 6

US-09-966-880A-7 (1-2818) x PHB3_HUMAN (1-382)
QY 92 TTGATGAACCGGAGGAAGTTCCTTTACCAATTCAAAATATCGCGTGGGCTAAGGTCGG 151
Db 192 LeuMetAspProAspThrPheAsnPheAsnAsnAspProLeuValLeuArgArg 211
QY 152 CGTGAGACCTACTGTGCTACGTAGTGAAGAGCGGTGACAGCGTGTACATCATCTTTTCACGTG 211
Db 212 ArgGlnThrTyLeuLeuGluValGluArgLeuAspAsnGlyThrTrpValLeuMet 231
QY 212 GAC-----TTTGGTTATCTTCGCAAT-----AAGAAC-----GGC 241
Db 232 AspGlnHisMetGlyPheLeuLeuGluAlaLysAsnLeuLeuGluCysGlyPheTyArg 251
QY 242 TCCACGCTGGAATTCCTCTTCCTCGCTACATCTCGGACTGGACCTAGACCTCGCCGC 301
Db 252 ArgHisAlaGluLeuArgPheLeuValProSerLeuGlnLeuAspProAlaGln 271
QY 302 TGCTACCGCGTCCACCTGGTTTCACCTCCCTGGAGCCCTCGTACGAC-----TGTGCCCGA 355
Db 272 IleTyArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGly 291

```

QY 356 CATGTGGCGGAGCTTTCTGCGAGGAGAACCCACCTCAGTCTGAGGATCTTCCACCGCGCC 415
 DB 292 GluValArgAlaPheLeuGlnGluAsnThrHisValArgLeuArgilePheAlaAArg 311
 QY 416 CTCCTACTTCTGTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTGCGCGCGCTG 466
 DB 312 IleTyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeu 326
 QY 467 CACCGCGCGGGTGCAATACCCATCATGACCTTCAAGAGATATTTTACTGCTGGGAT 526
 DB 327 ArgAspAlaGlyValGlnValSerIleMetThrTyrAspGluPheLeuTyrCysTrpasp 346
 QY 527 ACTTTGTGAGAAACCATGAAAGAACTTTCAAGCGCTGGGAAGGCTGCGATGAAATTC 586
 DB 347 ThrPheValTyrArgGlnGlyCysProPheGlnProTyrAspGlyLeuGluGluHisSer 366
 QY 587 GTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
 DB 367 GlnAlaLeuSerGlyArgLeuArgAlaIleLeu 377
 RESULT 13
 PHB1_HUMAN
 ID PHB1_HUMAN STANDARD; PRT; 199 AA.
 AC P31941; Q12807;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phorbolin 1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Epidermis;
 RX MEDLINE=99399284; PubMed=10469298;
 RA Madsen P.P., Anant S., Rasmussen H.H., Gromov P., Vorum H.,
 RA Dumanski J.P., Tommerup N., Collins J.E., Wright C.L., Dunham I.,
 RA Macginnitie A.J., Davidson N.O., Cellis J.E.;
 RT "Psoriasis up-regulated phorbolin-1 shares structural but not
 RT functional similarity to the mRNA-editing protein apobec-1.";
 RL J. Invest. Dermatol. 113:162-169(1999).
 RN [2]
 RP SEQUENCE OF 53-60; 112-121 AND 129-137.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=93162043; PubMed=1286667;
 RA Rasmussen H.H., van Damme J., Fuype M., Gesser B., Cellis J.E.,
 RA Vandekerckhove J.;
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel
 RT protein database of normal human epidermal keratinocytes.";
 RL Electrophoresis 13:960-969(1992).
 CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
 CC FAMILY. STRONG, TO APOLIPOPROTEIN B MRNA EDITING PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U03891; AAA03706.2;
 DR Aarhus/Ghent-2DPAGE; 2116; IEF.
 DR InterPro; IPR002125; dCMP/cytL_deam.
 DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.
 KW Hydrolase.
 SQ SEQUENCE 199 AA; 23012 MW; 42E99E0D7DF7AA14 CRC64;
 Alignment Scores:
 Pred. No.: 1.87e-24 Length: 199
 Score: 363.50 Matches: 83
 Percent Similarity: 58.12% Conservative: 28

Best Local Similarity: 43.46% Mismatches: 57
 Query Match: 7.35% Indels: 23
 DB: 1 Gaps: 7
 US-09-966-880A-7 (1-2818) x PHB1_HUMAN (1-199)
 QY 92 TTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGGCTAAGGCTCGG 151
 DB 12 LeuMetAspProHisIlePheThrSerAsnPheAsnAsn-----GlyIleGlyArg 28
 QY 152 COTGAGACCTACTGCTGTGCTAGTAGTAAGAGCGGTGACAGTGTATCATCTCTTTCAC 211
 DB 29 HisLysThrTyrLeuCysTyrGluValGluArgLeuAspAsnGlyThrSerValLysMet 48
 QY 212 GAC-----TTTGGTATCTTCGCRAT-----AGAAC-----GGC 241
 DB 49 AspGlnHisArgGlyPheLeuHisAsnGlnAlaLysAsnLeuLeuCysGlyPheTyrGly 68
 QY 242 TGCCACGCTGGAATTTCTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCCGC 301
 DB 69 ArgHisAlaGluLeuArgPheLeuAspLeuValProSerLeuGlnLeuAspProAlaGln 88
 QY 302 TGCTACCGCGCTCACCTGGTTACCTCTGAGCGCCCTGTACGAC-----TGTGCCCGA 355
 DB 89 IleTyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGly 108
 QY 356 CATGTGGCGGACTTTCTGCGAGGAGAACCCACCTCAGTCTGAGGATCTTCCACCGCGCC 415
 DB 109 GluValArgAlaPheLeuGlnGluAsnThrHisValArgLeuArgilePheAlaAArg 128
 QY 416 CTCCTACTTCTGTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTGCGCGCGCTG 466
 DB 129 IleTyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeu 143
 QY 467 CACCGCGCGGGTGCAATACCCATCATGACCTTCAAGAGATATTTTACTGCTGGAAT 526
 DB 144 ArgAspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheLysHisCysTrpasp 163
 QY 527 ACTTTGTGAGAAACCATGAAAGAACTTTCAAGCGCTGGGAAGGCTGCGATGAAATTC 586
 DB 164 ThrPheValAspHisGlnGlyCysProPheGlnProTyrAspGlyLeuAspGluHisSer 183
 QY 587 GTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
 DB 184 GlnAlaLeuSerGlyArgLeuArgAlaIleLeu 194
 RESULT 14
 ALU3_HUMAN
 ID ALU3_HUMAN STANDARD; PRT; 587 AA.
 AC P39190;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alu subfamily SBI sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert";
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:12:03 ; Search time 172.791 Seconds
(without alignments)
6720.711 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 4948
Sequence: 1 agagaaccatcataattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_n2p.model -DEV=xl
-Q=/cgn2_1/USPTO_spool/US09966880/runat_14062003_175524_10316/app_query.fasta_1.9493
-DB=SPTRMBL_21 -FMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880.BCGN.1.1.978 @runat_14062003_175524_10316 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1086	21.9	198 4 Q9GZX7	Q9gzx7 homo sapien

2	1008	20.4	198	11	Q9WVE0	Q9wve0 mus musculus
3	403.5	8.2	171	4	Q9H728	Q9h728 homo sapien
4	390	7.9	384	4	Q9HC16	Q9hc16 homo sapien
5	389	7.9	136	4	Q96NR6	Q96nr6 homo sapien
6	381.5	7.7	239	4	Q9NX17	Q9nx17 homo sapien
7	370	7.5	118	4	Q9H387	Q9h387 homo sapien
8	369.5	7.5	204	4	Q9Y555	Q9y555 homo sapien
9	369	7.5	122	6	Q9BGW3	Q9bgw3 macaca fasc
10	367.5	7.4	190	4	Q96F12	Q96f12 homo sapien
11	360	7.3	133	4	Q96JRS	Q96jrs homo sapien
12	357	7.2	127	4	Q9H743	Q9h743 homo sapien
13	356.5	7.2	152	4	Q9NX85	Q9nx85 homo sapien
14	354.5	7.2	294	4	Q9NRW3	Q9nrw3 homo sapien
15	350	7.1	429	11	Q99J72	Q99j72 mus musculus
16	341.5	6.9	386	4	Q96AK3	Q96ak3 homo sapien
17	339	6.9	115	6	Q9N083	Q9n083 macaca fasc
18	336	6.8	123	4	Q9HAD8	Q9had8 homo sapien
19	321.5	6.5	238	4	Q8WTZ3	Q8wtz3 homo sapien
20	320.5	6.5	666	4	P78525	P78525 homo sapien
21	319	6.5	130	4	Q9HBS7	Q9hbs7 homo sapien
22	317	6.4	169	4	Q9H397	Q9h397 homo sapien
23	315.5	6.4	139	4	Q9Y4V1	Q9y4v1 homo sapien
24	303	6.1	162	4	Q9BYA5	Q9bya5 homo sapien
25	303	6.1	165	4	Q96LS9	Q96ls9 homo sapien
26	303	6.1	375	4	O60448	O60448 homo sapien
27	298.5	6.0	118	4	Q9P195	Q9p195 homo sapien
28	294.5	6.0	368	4	O15662	O15662 homo sapien
29	292	5.9	214	4	Q9NRE8	Q9nre8 homo sapien
30	289	5.9	111	4	Q9H3C0	Q9h3c0 homo sapien
31	287	5.8	231	4	Q9NR08	Q9nr08 homo sapien
32	284	5.7	121	4	Q96N97	Q96n97 homo sapien
33	279	5.7	91	4	Q9BYA9	Q9bya9 homo sapien
34	279	5.7	232	4	Q9H5R3	Q9h5r3 homo sapien
35	278	5.6	224	11	Q9WV35	Q9wv35 mus musculus
36	277	5.6	90	4	Q96IG1	Q96ig1 homo sapien
37	275	5.6	535	4	Q96EB1	Q96eb1 homo sapien
38	274.5	5.6	151	4	Q9HA67	Q9ha67 homo sapien
39	274	5.5	224	4	Q9Y235	Q9y235 homo sapien
40	272	5.5	139	4	Q9BVD9	Q9bvd9 homo sapien
41	272	5.5	163	4	Q96MM0	Q96mm0 homo sapien
42	270.5	5.5	120	4	Q9H6G8	Q9h6g8 homo sapien
43	270.5	5.5	121	4	Q9Y553	Q9y553 homo sapien
44	263	5.3	109	4	Q8WZ39	Q8wz39 homo sapien
45	261	5.3	83	4	Q96ID7	Q96id7 homo sapien

ALIGNMENTS

RESULT 1

Q9GZX7 ID - Q9GZX7 PRELIMINARY; PRT; 198 AA.
AC Q9GZX7;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Activation-induced cytidine deaminase.
GN AID.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20408890; PubMed=10950930;
RA Muto T., Muramatsu M., Tanikawa K., Kinoshita K., Honjo T.;
RT "Isolation, tissue distribution and chromosomal localization of the
RT human activation-induced cytidine deaminase (AID) gene.";
RL Genomics 68:85-88(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20460541; PubMed=11007475;
RA Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,
RA Catalan N., Forvelle M., Dufourcq-Lagelouse R., Gennery A.,

RA Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N.,
RA Muramatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A.,
RA Durandy A.;
RT "Activation-induced cytidine deaminase (AID) deficiency causes the
RT autosomal recessive form of the Hyper-IgM syndrome (HIGM2).";
RL Cell 102:563-575(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=B-CELL;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB040431; BAB12721.1; -;
DR EMBL; AB040430; BAB12720.1; -;
DR EMBL; BC006296; AAH06296.1; -;
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 198 AA; 23953 MW; 3C27BB143DB184A9 CRC64;

Alignment Scores:
Pred. No.: 2,65e-105 Length: 198
Score: 1086.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.95% Indels: 0
Gaps: 4

US-09-966-880A-7 (1-2818) x Q9GZX7 (1-198)

QY 80 ATGCAGACCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCTGG 139
Db 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGGCGTGACGTGTCTACA 199
Db 21 AlaLysGlyArgargGluThrTyrLeuCysTyrValValLysArgargSerpAlaThr 40
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 259
Db 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
QY 260 TTCCTCCGCTACATCTCGGACTGGACCTAGACCTGGCGCTGCTACCGCGTCACCTGG 319
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 320 TTCACCTCTCGGAGCCCTCTACGACTGTGCCCGCATGTGCCGACTTCTCGCGAGGG 379
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
QY 380 AACCCCACTCAGCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGGACCGCAAG 439
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 440 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 499
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
QY 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 559
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
QY 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCGATCCTT 619
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
QY 620 TTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTT 673
Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2
Q9WVE0 PRELIMINARY; PRT; 198 AA.
AC Q9WVE0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Activation-induced cytidine deaminase.
GN ATCDA OR AID.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303612; PubMed=10373455;
RA Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,
RA Davidson N.O., Honjo T.;
RT "Specific expression of activation-induced cytidine deaminase (AID), a
RT novel member of the RNA-editing deaminase family in germinal center B
RT cells.";
RL J. Biol. Chem. 274:18470-18476(1999).
DR EMBL; AF132979; AAD41793.1; -;
DR MGD; MGI:1342279; Aicda.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BEB2 CRC64;

Alignment Scores:
Pred. No.: 4,51e-97 Length: 198
Score: 1008.00 Matches: 183
Percent Similarity: 95.94% Conservative: 6
Best Local Similarity: 92.89% Mismatches: 8
Query Match: 20.37% Indels: 0
Gaps: 11

US-09-966-880A-7 (1-2818) x Q9WVE0 (1-198)

QY 80 ATGCAGACCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCTGG 139
Db 1 MetAspSerLeuLeuMetLysGlnLysLysPheLeuTyrHisPheLysAsnValArgTrp 20
QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGGCGTGACGTGTCTACA 199
Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgArgSerpAlaThr 40
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 259
Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeuLeu 60
QY 260 TTCCTCCGCTACATCTCGGACTGGACCTAGACCTGGCGCTGCTACCGCGTCACCTGG 319
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 320 TTCACCTCTCGGAGCCCTCTACGACTGTGCCCGCATGTGCCGACTTCTTCGCGAGGG 379
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
QY 380 AACCCCACTCAGCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGGACCGCAAG 439
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 440 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 499
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140
QY 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 559
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160
QY 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCGATCCTT 619
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180
QY 620 TTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGA 670
Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197


```

Q96NR6
ID Q96NR6 PRELIMINARY; PRT; 136 AA.
AC Q96NR6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE CDNA FLJ30278 fis, clone BRACE2002755.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054840; BAB70813.1;
SQ SEQUENCE 136 AA; 14980 MW; D2336B649A110163 CRC64;

Alignment Scores:
Pred. No.: 8 61e-32 Length: 136
Score: 389.00 Matches: 73
Percent Similarity: 80.20% Conservative: 8
Best Local Similarity: 72.28% Mismatches: 19
Query Match: 7.86% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x Q96NR6 (1-136)
QY 1830 GCGCGGGGTGGTCCACGCTGTAATCCAGCACTTTGGAGCGGAGCCGGCGGA 1889
Db 30 GlyLeuAlaGlnTrpLeuThrProValIleProAlaLeuTrpGluThrGluAlaGlyGly 49
QY 1890 TCACCTGTGGTCAGGAGTTTCAGACCCAGCCCTGGCCACATGGCAAAACCCCGTGTACT 1949
Db 50 SerLeuGluValArgSerArgProAlaTrpProThrTrpArgAsnProIleSerThr 69
QY 1950 CAAATGCAAAATAGCGGCTGTAGCAGCACCCTGTAACTCCAGCTACTTGGGAG 2009
Db 70 LysAsnThrLysIleSerGlnThrTrpTrpArgMetProValValLeuAlaThrTrpGlu 89
QY 2010 GCTGAGCAGAGAAATCGTTGAACCCAGGAGTGGAGTTCAGTAAGCTGATCGTG 2069
Db 90 AlaGluAlaGlyGluSer-ValAsnLeuGlyGlyArgGlyCysSerGluLeuArgGlyCy 109
QY 2070 CCGTTGCACTCCAGCCCTGGCGGACAGCAAGCACTGTCTCAGAAAAAATAAAAAA 2129
Db 109 SARGCyThrProAlaTrpAlaThrArgAlaLysLeuHisLeuArgLysAsnLysAsnLy 129
QY 2130 A 2130
Db 129 s 129

RESULT 6
Q9NX17
ID Q9NX17 PRELIMINARY; PRT; 239 AA.
AC Q9NX17;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE CDNA FLJ20489 fis, clone KAT08285.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

Q9H387
ID Q9H387 PRELIMINARY; PRT; 118 AA.
AC Q9H387;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE PRO2550
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
RT "Functional prediction of the coding sequences of 75 new genes deduced
RL by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130089; AAG35515.1;
SQ SEQUENCE 118 AA; 13257 MW; 94688870CAC8760D CRC64;

Alignment Scores:
Pred. No.: 8 39e-30 Length: 118
Score: 370.00 Matches: 77

Q9H387
ID Q9H387 PRELIMINARY; PRT; 118 AA.
AC Q9H387;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE PRO2550
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
RT "Functional prediction of the coding sequences of 75 new genes deduced
RL by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130089; AAG35515.1;
SQ SEQUENCE 118 AA; 13257 MW; 94688870CAC8760D CRC64;

Alignment Scores:
Pred. No.: 8 39e-30 Length: 118
Score: 370.00 Matches: 77

```



```

Db      44  uLysGlnPheSerCysLeuSerLeuProSerSerTrpAsnTyrArgCysLeuProProH1 64
QY      1971  CTTGGCTAAATTTGGATTTGAGTACAGAGGGGTTTCCCATGTTGGCCAGCGTGTC 1912
Db      64  sLeuAlaLysPheCysIlePheSer-Argasp-----TyrValGlyGlnAlaGlyL 81
QY      1911  TCAAACTCTCTGACACAGAGGTATCCGCCCGCTCGGCTCCCAAGTGTCTGGGATTACAG 1852
Db      81  euGluLeuLeuThrSerGlyAsnProLeuAlaSerAlaSerGlnSerAlaGlyIleThrG 101
QY      1851  GCCTGAGCACACAGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1811
Db      101  lyMetSerHisHisSerArgProGlnMetPheLeuGluPhe 114

RESULT 10
Q96F12
ID      Q96F12      PRELIMINARY;      PRT;      190 AA.
AC      Q96F12;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE      Similar to APOBEC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=UTERUS;
RA      Strausberg R.;
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC011739; AAH11739.1; -.
DR      InterPro: IPR002125; dCMP/cytL_deam.
DR      PROSITE: PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ      SEQUENCE 190 AA; 22827 MW; DA0584EF75C91CF0 CRC64;

```

```

Alignment Scores:
Pred. No.:      1,75e-29      Length:      190
Score:          367.50      Matches:      79
Percent Similarity: 58.52%      Conservative: 24
Best Local Similarity: 44.89%      Mismatches: 64
Query Match:      7.43%      Indels: 9
DB:              4      Gaps: 4

```

```

US-09-966-880A-7 (1-2818) x Q96F12 (1-190)
QY      110  TTTCTTTTACCAATTCAAAATGTCCTGGCTAAGGTCGGGTGAGACCTACCTGTGC 169
Db      17  PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAspGluThrTrpLeuCys 36
QY      170  TAGCTAGTGAAG-----AGCGGTGACAGTGTACATCTCTTTTCACCTGGACTTTGGTTAT 223
Db      37  PheThrValCluGlyIleLysArgArgSerValValSerTrpLysThr-----GlyVal 54
QY      224  CTTCCGAAT-----AGAACGGCTGCCAGTGAATGTCTCTCCCTCCGCTAC 271
Db      55  PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY      272  ATCTCGGACTGGGACCTAGACCTCGCGCTGCTACCGCTACCTGGTTCACCTCCTCG 331
Db      75  PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrSerTrp 94
QY      332  AGCCCTGCTAGCATGTGCCCGACATGTGGCGCACTTTCTGGAGGGAACCCCAACCTC 391
Db      95  SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY      392  AGTCTCAGATCTTACCGCGGCCCTCTACTTCTGTGAGACCGCAGGCTGAGCCCGAG 451
Db      115  AsnLeuThrIlePheThrAlaArgLeuTyrPheGln---TyrProCysTyrGlnGlu 133
QY      452  GGGCTCGCGGGCTGCACCGCGCGGGTGCRAATAGCCATCATGACCTTCAAAGATTAT 511

```

A

```

Db      134  GlyLeuArgSerLeuSerGlnGlyValAlaValGluLeuMetAspTyrGluAspPhe 153
QY      512  TTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTTTCAAAGCTCGGAAGGG 571
Db      154  LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
QY      572  CTCATGAAATTAACAGTTCTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db      174  LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 11
Q96JR5
ID      Q96JR5      PRELIMINARY;      PRT;      133 AA.
AC      Q96JR5;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      PRO0764.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RX      MEDLINE=21376145; PubMed=11483580;
RA      Yu Y., Zhang C., Zhou G., Wu S., Ou X., Wei H., Xing G., Dong C.,
RA      Zhai Y., Wan J., Ouyang S., Li L., Zhang S., Zhou K., Zhang Y., Wu C.,
RA      He F.;
RT      "Gene Expression Profiling in Human Fetal Liver and Identification of
RT      Tissue- and Developmental-Stage-Specific Genes through Compiled
RT      Expression Profiles and Efficient Cloning of Full-Length cDNAs.";
RL      Genome Res. 11:1392-1403(2001).
DR      EMBL; AF305818; AAK55521.1; -.
SQ      SEQUENCE 133 AA; 14414 MW; C9318E532FB139F CRC64;

```

```

Alignment Scores:
Pred. No.:      9.83e-29      Length:      133
Score:          360.00      Matches:      68
Percent Similarity: 82.98%      Conservative: 10
Best Local Similarity: 72.34%      Mismatches: 16
Query Match:      7.29%      Indels: 0
DB:              4      Gaps: 0

```

```

US-09-966-880A-7 (1-2818) x Q96JR5 (1-133)
QY      2071  GGCACATCTACGCTTACTCCAACTCCACCTCTCGGTTCAGCGCATTCCTCGCTCA 2012
Db      29  GlySerLeuGlnThrHisCysAsnIleHisLeuProGlySerSerAspSerProAlaSer 48
QY      2011  GCCTCCCAAGTAGCTGGGATTACAGCTGCTGTACAGCGCTGGCTAATTTTGCATTT 1952
Db      49  AlaSerTrpValAlaGlyIleThrGlyThrArgHisHisThrTrpLeuIlePheValPhe 68
QY      1951  TGAGTACAGACGGGGTTTGGCCATGTTGGCAGGCTGTCTCAAACTCTGACACAGGT 1892
Db      69  LeuValValThrGlyPheHisHisValGlyGlnAlaGlyLeuLysLeuLeuThrSerGly 88
QY      1891  GATCCCGCGCGCTCGGCTCCCAAACTGTGGGATTACAGCGCTGAGCCACCGCCCG 1832
Db      89  AspProProSerSerSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaGln 108
QY      1831  CCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1790
Db      109  ProTyrSerCysIlePhePhePhePhePhePheGluMetGlu 122

```

```

RESULT 12
Q9H743
ID      Q9H743      PRELIMINARY;      PRT;      127 AA.
AC      Q9H743;
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT      01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

```


[illegible]

```
QY 110 TTTCTTTACCAATTCAAAATGTCGCTGGCTGAAGGTCGCGGTGAGACCTTACCTGTGC 169
Db 17 PheTyrPheGlnPheLeuSerGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 170 TACGTAGTGAAG-----AGCGGTGACAGTGTACATCTTTTCTACTGGACTTTGGTTAT 223
Db 37 PheThrValGluGlyIleLysArgSerValValSerTrpLysThr-----GlyVal 54
QY 224 CTTCGCAAT-----AAGACGGCTGCCACGTGGAATTCCTCTCTCGCTAC 271
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY 272 ATCTCGGACTGGACCTAGACCTGGCGCTGTCTACCGCGTACCTGGTTCACCTCTGG 331
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94
QY 332 AGCCCTCTAGACTGTGCGCGACATGTGCGCGACTTCTCCGAGGAAACCCCACTC 391
Db 95 SerProCysProAspCysAlaGlyValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 392 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGACGCTGAGCCGCG 451
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY 452 GGGCTGGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTAT 511
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153
QY 512 TTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGACTTTTCAAGCTGCGGAAGG 571
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpGluGly 173
QY 572 CTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCTTTTGGCCCTGTAT 631
Db 174 Ile---LysAsnGlnLeuSerThrSerGluLysLysAlaThrGlyGluSerProValArg 192
QY 632 GAGGTTGATGACTTACGAGAGCGATTCGTTACTTTGGGACTTTGATAGCAACTTCCAG 691
Db 193 -----GlyLeuProGly 196
QY 692 ATGTCACACAGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCTCAAGT 751
Db 197 ---ProHisGlyLeuSerProLeuAlaSerCysSerCysThrGlyLeuProSerThr 215
QY 752 CTT-----CTCTGTTT-----TATPCT 769
Db 216 LeuAspProLeuCysPheCysLeuValIleLeuSerProSerTrpProGlnGlyHisSer 235
QY 770 TCAACTCTCACTTCTTGTAGAGTTTACAGAAAAAATATT 808
Db 236 ThrValLeuProCysLeuThrAlaSerSerLeuPhe 248
```

RESULT 15

```
Q99J72 PRELIMINARY; PRT; 429 AA.
AC Q99J72;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 51.0 kDa protein.
GN BC003314.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003314; AA003314.1;
DR MGD; MGI:193311; BC003314.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_2.
```

```
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 51017 MW; BE44D01380AD7F6E CRC64;
```

Alignment Scores:

```
Pred. No.: 1.52e-27 Length: 429
Score: 350.00 Matches: 81
Percent Similarity: 49.39% Conservative: 40
Best Local Similarity: 33.06% Mismatches: 84
Query Match: 7.07% Indels: 40
Db: 11 Gaps: 6
```

US-09-966-880A-7 (1-2818) x Q99J72 (1-429)

```
QY 92 TTGTATGACCCGGAGGAGAGTTCCTTTTACCAATTCAAAATGTCGCGTAAAGGTGCG 151
Db 20 LeuIleSerGlnGluThrPheLysPheHisPheLysAsnLeuArgTyrAlaIleAspArg 39
QY 152 CGTGAGACCTACCTGTCTAGTAGTAAGAGGCGGTGACAGTGTACATCTCTTCTACTG 211
Db 40 LysAspThrPheLeuCysTyrGluValThrArgLysAspCysAspSerProValSerLeu 59
QY 212 GACTTTGGTTATCTTCGCAATAGACGGGTGCCACGTGGAATTCCTTCTCTCCGCTAC 271
Db 60 HisHisGlyValPheLysAsnLysAspAsnIleHisAlaGluIleCysPheLeuTyrTrp 79
QY 272 ATCTCGGAC-----TGGGACCTAGACCTGGCGCTGCTACCGCGCTCACCTGGTTC 322
Db 80 PheHisLysValLeuLysValLeuSerProArgGluGluPheLysIleThrTrpTrp 99
QY 323 ACCTCTGGAGCCCTCTCTACGACTGTGCCCGACATGTGCCGAGCTTTCTCGGAGGCAAC 382
Db 100 MetSerTrpSerProCysPheGluCysAlaGluValLeuArgPheLeuAlaThrHis 119
QY 383 CCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCT 442
Db 120 HisAsnLeuSerLeuAspIlePheSerSerArgLeuTyrAsnIleArgAspProGluAsn 139
QY 443 GAGCCCGAGGGGTGCGCGGCTGACCGCGCGGCTGCAATAGCATCATGACCTTC 502
Db 140 Gln---GlnAsnLeuCysArgLeuValGlnGluGlyAlaGlnValAlaAlaMetAspLeu 158
QY 503 AAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGACTTTTCAAGGCC 562
Db 159 TyrGluPheLysLysCysTrpLysLysPheValAspAsnGlyArgArgPheArgPro 178
QY 563 TGGGAAGGGGTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 622
Db 179 TrpLysLeuLeuLeuThrAsnPheArgTyrGlnAspSerLysLeuGlnGluIleLeuArg 198
QY 623 CCCTCTGATGAGTTGATGACTTACGAGACGCACTTTCGTTACTTTGGGACTTTTGATAC 682
Db 199 ProCysTyr----- 201
QY 683 CTTCGA-----GGAATGTCACACAGATGAATATCTCTGTGGAACACAGTGGATA 733
Db 202 IleProValProSerSerSerSerThrLeuSerAsnIleCys-----LeuThr 218
QY 734 AAAACAGCTCTTCAAGTCTTCTCTCT----- 760
Db 219 LysGlyLeuProGluThrArgPheCysValGluGlyArgArgValHisLeuLeuSerGlu 238
QY 761 -----TTTTATTCT 769
Db 239 GluGluPheTyrSer 243
```

Search completed: June 14, 2003, 18:47:39
Job time : 181.791 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:13:03 ; Search time 27.1417 Seconds
(without alignments)
6109.704 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 4948
Sequence: 1 agagaaccatcataattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US0966880/runat_14062003_175525_10370/app_query.fasta_1.9493
-DB=issued_Patents_AA -Qfwt=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosun62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US0966880.cgn1_1_139.eunat_14062003_175525_10370 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	7.5	190	1	US-08-816-241-1
2	369.5	7.5	190	3	US-09-128-395-1
3	332	6.7	132	4	US-09-605-785-573
4	303	6.1	375	2	US-08-454-557C-121
5	303	6.1	375	2	US-08-340-426D-121
6	303	6.1	375	2	US-08-450-673C-121
7	295	6.0	397	5	PCT-US95-17111A-121
8	287	5.8	1079	3	US-09-058-489-22
9	286	5.8	99	4	US-09-288-143-168
10	278	5.6	110	4	US-09-227-357-193
11	271	5.5	222	1	US-08-687-895-1
12	271	5.5	222	2	US-09-040-482-1

c	13	248	5.0	122	4	US-09-227-357-285	Sequence 285, App
	14	245	5.0	116	1	US-08-687-895-3	Sequence 3, Appli
	15	245	5.0	116	1	US-08-816-241-3	Sequence 3, Appli
	16	245	5.0	116	2	US-09-040-482-3	Sequence 3, Appli
	17	245	5.0	116	3	US-09-128-395-3	Sequence 3, Appli
c	18	225.5	4.6	90	4	US-09-227-357-171	Sequence 171, App
	19	221	4.5	236	1	US-08-687-895-4	Sequence 4, Appli
	20	221	4.5	236	1	US-08-816-241-4	Sequence 4, Appli
	21	221	4.5	236	2	US-09-040-482-4	Sequence 4, Appli
	22	221	4.5	236	3	US-09-128-395-4	Sequence 4, Appli
c	23	218.5	4.4	56	4	US-09-227-357-577	Sequence 577, App
	24	214.5	4.3	236	1	US-08-158-682A-4	Sequence 4, Appli
	25	199.5	4.0	229	1	US-08-158-682A-2	Sequence 2, Appli
	26	199.5	4.0	229	1	US-08-015-203-2	Sequence 2, Appli
	27	199.5	4.0	229	1	US-08-687-895-5	Sequence 5, Appli
	28	199.5	4.0	229	1	US-08-816-241-5	Sequence 5, Appli
	29	199.5	4.0	229	2	US-09-040-482-5	Sequence 5, Appli
	30	199.5	4.0	229	3	US-09-128-395-5	Sequence 5, Appli
	31	191.5	3.9	76	4	US-09-605-785-575	Sequence 575, App
c	32	165.5	3.4	500	4	US-09-265-630-13	Sequence 13, Appli
	33	160	3.2	72	4	US-09-227-357-655	Sequence 655, App
c	34	157	3.2	368	4	US-09-265-630-11	Sequence 11, Appli
c	35	156.5	3.2	579	3	US-08-704-711A-1	Sequence 1, Appli
	36	145.5	3.2	579	4	US-09-521-220-1	Sequence 1, Appli
	37	145.5	2.9	52	4	US-09-227-357-537	Sequence 537, App
	38	134	2.7	447	1	US-08-365-689-1	Sequence 1, Appli
	39	134	2.7	447	1	US-07-747-781-1	Sequence 1, Appli
	40	134	2.7	447	1	US-08-145-138A-1	Sequence 1, Appli
	41	134	2.7	447	5	PCT-US92-03088-1	Sequence 1, Appli
	42	134	2.7	447	5	PCT-US93-03027-5	Sequence 5, Appli
c	43	128	2.6	47	4	US-09-227-357-656	Sequence 656, App
c	44	120	2.4	48	3	US-08-951-200A-1	Sequence 1, Appli
c	45	120	2.4	619	3	US-09-156-253-48	Sequence 48, Appli

ALIGNMENTS

RESULT 1
US-08-816-241-1
: Sequence 1, Application US/08816241
: Patent No. 5804185
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Goli, Surya K.
: TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/816,241
: FILING DATE: Filed Herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0239 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 1:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
;

```

Alignment Scores:		
Pred. No.:	2.94e-32	Length:
Score:	369.50	Matches:
Percent Similarity:	58.52%	Conservative:
Best Local Similarity:	44.89%	Mismatches:
Query Match:	7.47%	Indels:
DB:	1	Gaps:

US-09-966-880A-7 (1-2818) x US-08-816-241-1 (1-190)

110	QY	TTTCTTTACCAATTCAAAAATGTC	CCGTGGCTAAAGGTCGGCTGAGACCTACCTGTGC	169	
	Db				
17	PhetYrPheGlnPheLysAsnLeuT	riPgluAlaAsnAspArgAsnGluThr	TrpLeuCys	36	
	Db				
170	QY	TACGTAGTGAAG-----AGCGCG	CACAGTGCTACATCCCTTTTCAC	TGGACITTTGGTTAT	223
	Db				
37	PheThrValGluGlyIleLysArg	SerValValSerTrpLysThr	-----GlyVal	54	
	Db				
224	QY	CTTGCAAT-----AAGACGCG	TGCCAGTGGGAATGCTCTTCCTCCGCTAC	271	
	Db				
55	PheArgAsnGlnValAspSerGlu	ThrHisCysHisAlaGluArgCys	PheLeuSerTrp	74	
	QY	ATCTCGGACTGGGACCTACACCT	GGCGCGTGTACCGGTACCTGGTTCACCTCCTCGG	331	
	Db				
75	PheCysAspAlleLeuSerPro	AsnThrLysTyrGlnValThrTrp	TrpSerTrp	94	
	Db				
332	QY	AGCCCTGTCTACGACTGTGCCCC	GACATGTGGCGGACTTTCTCGCGAGGAAACCCCAACCTC	391	
	Db				
95	SerProCysProAspCysAlaGly	GluValAlaIaGluPheLeuAla	ArgHisSerAsnVal	114	
	QY	AGTCTGAGGACTTTCACCGCGC	CTCTACTTCTGTGAGGACCGCAAGCTCAGCCCGAG	451	
	Db				
115	AsnLeuThrIlePheThrAlaArg	LeuTyrTyrPheGln-----TyrPro	CysTyrGlnGlu	133	
	QY	GGGTCTGGCGGTGCACCGCGCG	GGTGCAAATAGCCATCATGACCTTCAAGATTAT	511	
	Db				
134	GlyLeuArgSerLeuSerGlnGlu	GlyValAlaValGluIleMetAsp	TyrGluAspPhe	153	
	QY	TTTTTACTGCTGGAATACTTTGT	AGAAAACCTTTCAAAGCCCTGGCAAGG	571	
	Db				
154	LysTyrCysTrpGluAsnPheVal	TyrAsnAspAsnGluProPhe	LysProTrpLysGly	173	
	QY	CTGCATCAAAATTCAGTTCGTCT	CTCCAGACAGCTTCGGCGCATCCTT	619	
	Db				
174	LeuLysThrAsnPheArgLeuLeu	LysArgLeuArgGluSerLeu	189		
	Db				

RESULT 2

US-09-128-395-1
: Sequence 1, Application US/09128395
: Patent No. 6087108
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Goli, Surya K.
: TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:

Db 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 3

US-09-605-785-573

; Sequence 573, Application US/09605785

; Patent No. 6321716

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqul

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skelky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C16

; CURRENT APPLICATION NUMBER: US/09/605,785

; CURRENT FILING DATE: 2000-06-27

; NUMBER OF SEQ ID NOS: 835

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 573

; LENGTH: 132

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-605-785-573

Alignment Scores:

Pred. No.: 3 51e-28 Length: 132

Score: 332.00 Matches: 65

Percent Similarity: 71.43% Conservative: 5

Best Local Similarity: 66.33% Mismatches: 28

Query Match: 6.73% Indels: 0

DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-605-785-573 (1-132)

QY 2123 TTTTCTTCTGAGACAGAGTCTTGCTGTGCGCCAGCGCTGAGTGCAACGGCACGAT 2064

Db 28 PhePhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGlnTrpHisAsp 47

QY 2063 CTCAGCTTACTGCAACCTCCACCTCTCTGGGTTCAACGGATTCTCTCGCTCAGCTCCCA 2004

Db 48 LeuSerSerLeuGlnProLeuProHisArgPheLeuGlnPheSerCysLeuSerLeuPro 67

QY 2003 AGTAGCTGGATTACAGGTGCGCTGTACACACGCTCGCTAAATTTTTCATTTTGACTACA 1944

Db 68 HisSerTrpAspHisArgTrpAlaProProHisLeuAlaAsnPheCysSerPheSerArg 87

QY 1943 GACGGGGTTTGCATGTTGGCCAGGCTGGTCTCAAACTCCTGACACAGGTGATCCGCC 1884

Db 88 AspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGlnGlnSerAla 107

QY 1883 CGGCTCGGCTCCCAAGTGTGGGATTACAGCGGTGAGCCACCGCCCGCCGCC 1830

Db 108 CysLeuGlyLeuProLysCysTrpGlyTyraGHisLysProHisProAla 125

RESULT 4

US-08-454-557C-121

; Sequence 121, Application US/08454557C

; Patent No. 5830670

; GENERAL INFORMATION:

; APPLICANT: de la Monte, Suzanne

; APPLICANT: Wands, Jack R.

; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

; NUMBER OF SEQUENCES: 121

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/454,557C

; FILING DATE: 30-MAY-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, Steven R.

; REGISTRATION NUMBER: 36,203

; REFERENCE/DOCKET NUMBER: 0609.3840003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 121:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 375 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-454-557C-121

Alignment Scores:

Pred. No.: 1 05e-24 Length: 375

Score: 303.00 Matches: 90

Percent Similarity: 47.14% Conservative: 9

Best Local Similarity: 42.86% Mismatches: 41

Query Match: 6.14% Indels: 71

DB: 2 Gaps: 6

US-09-966-880A-7 (1-2818) x US-08-454-557C-121 (1-375)

QY 2098 CTCCTGTCGCCAGGCTGGAGTCCAAACGGCACGATCTCAGCTTACTGCAACCTCCACCTC 2039

Db 5 LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArgLeu 24

QY 2038 CTGGGTTCAAGCGATTCTCTGCTCAGCTCCCAAGTAGCTGGGATTACAGGTGCGCTGC 1979

Db 25 ProGlySerSerAspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCys 44

QY 1978 TACCACGCTGGCTAAAT---TTTGATTTTGATACAGACGGGGTTTGGCCATGTGGC 1922

Db 45 ThrHisAlaArgLeuIleLeuTyPhePheLeuValGluMetGluPheLeuHisValGly 64

QY 1921 CAGGCTGGCTCAAACTCCTGACACAGGTGATCCGCCGG----- 1881

Db 65 GlnAlaGlyLeuGluLeuProThrSerAspProSerValSerAlaSerGlnSerAla 84

QY 1881 ----- 1881

Db 85 ArgTyArgThrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArg 104

QY 1880 -----CTCGGCTC 1872

Db 105 ValSerLeuMetCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeu 124

QY 1871 CCNAAGTGTGGATTACAGGCGTGAGCCACCGCCCGCCCTCTCTCTTTCTTTT 1812

Db 125 ProLysCysTrpAspTyArgArg--AlaAlaValProGly-----LeuPheI 140

QY 1811 TTTTCTTTTGTGAGACAGGATCTTGTATGTTGCCA----- 1771
 Db 140 leuPheLeuArgHisArg-----CysProThrLeuThrGlnAspGluV 156
 QY 1770 --GGCTGG-----CCTTGAACACCTGTGTCAAGCCATCTCTCCC 1734
 Db 156 a1GlnTrpCysAspHisSerLeuGlnProSerThrProGluIleLeys-HisProPro 175
 QY 1733 TCCTCAACCTCCCAAGCAGCTAGAACTATGAGCGTGTGCCACGGTTCGCACCTTTATCAT 1674
 Db 176 AlaSerAlaSerGlnValAlaGlyThrLysAspMetHisLysThrTrpLeu-IlePh 195
 QY 1673 TCTGTTTGTACAACTCTCTT 1652
 Db 195 e1lePheIlePheAsnPhLeu 202

RESULT 5
 US-08-340-426D-121
 ; Sequence 121, Application US/08340426D
 ; Patent No. 5948634
 ; GENERAL INFORMATION:
 ; APPLICANT: de la Monte, Suzanne
 ; TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection
 ; TITLE OF INVENTION: of Alzheimer's Disease
 ; NUMBER OF SEQUENCES: 121
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005-3934

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 APPLICATION DATA:
 FILING DATE: 14-NOV-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, Steven R.
 REGISTRATION NUMBER: 36,203
 REFERENCE/DOCKET NUMBER: 0609.3840002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 121:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 375 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-340-426D-121
 Alignment Scores:
 Pred. No.: 1,05e-24 Length: 375
 Score: 303.00 Matches: 90
 Percent Similarity: 47.14% Conservative: 9
 Best Local Similarity: 42.86% Mismatches: 41
 Query Match: 6.14% Indels: 71
 DB: 2 Gaps: 6

US-09-966-880A-7 (1-2818) x US-08-340-426D-121 (1-375)
 QY 2098 CTCCTTCGCCAGCGTGGAGTGCACGCGCAGCTACTGCAACTCCACCTC 2039
 Db 5 LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArgLeu 24
 QY 2038 CTGGGTTCAAGGATTCCTCTCCCTCAGCTCCCGAGTAGCTGGGATTACAGTGCCTGC 1979

Db 25 ProGlySerSerAspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCys 44
 QY 1978 TACCAGCCCTGGCTAATT---TTTGCAATTTGAGTACAGACGGGTTTTCCTCATGTTGCG 1922
 Db 45 ThrHisAlaArgLeuIleLeuThrPhePheLeuValGluMetGluPheLeuHisValGly 64
 QY 1921 CAGGCTGGTCTCAAACTCTCTCACCACAGGTGATCCGCCCGG----- 1881
 Db 65 GlnAlaGlyLeuGluLeuProThrSerAspProSerValSerAlaSerGlnSerAla 84
 QY 1881 ----- 1881
 Db 85 ArgTyrArgThrGlyHisHisAlaArgLeuCysLeuAlaAsnPhCysGlyArgAsnArg 104
 QY 1880 -----CTCGGCGCTC 1872
 Db 105 ValSerLeuMetCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeu 124
 QY 1871 CCAAAGTCTGGGATTACAGCGTGAGCCACCCAGCCGCGCTCTCTCTCTCTCTCTCTCTCT 1812
 Db 125 ProLysCysTrpAspTyrArgArg--AlaAlaValProGly-----LeuPheI 140
 QY 1811 TTTTCTTTTCTGAGACAGCAGGATCTGTATGTTGCCCA----- 1771
 Db 140 leuPhePheLeuArgHisArg-----CysProThrLeuThrGlnAspGluV 156
 QY 1770 --GGCTGG-----CCTTGAACACCTGTGTCAAGCCATCTCTCCC 1734
 Db 156 a1GlnTrpCysAspHisSerSerLeuGlnProSerThrProGluIleLeys-HisProPro 175
 QY 1733 TCCTCAACCTCCCAAGCAGCTAGAACTATGAGCGTGTGCCACGGTTCGCACCTTTATCAT 1674
 Db 176 AlaSerAlaSerGlnValAlaGlyThrLysAspMetHisLysThrTrpLeu-IlePh 195
 QY 1673 TCTGTTTGTCAAACTCTCTT 1652
 Db 195 e1lePheIlePheAsnPhLeu 202

RESULT 6
 US-08-450-673C-121
 ; Sequence 121, Application US/08450673C
 ; Patent No. 5948888
 ; GENERAL INFORMATION:
 ; APPLICANT: de la Monte, Suzanne
 ; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
 ; TITLE OF INVENTION: of Alzheimer's Disease
 ; NUMBER OF SEQUENCES: 121
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Avenue, Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005-3934

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 FILING DATE: 30-MAY-1995
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Ludwig, Steven R.
 REGISTRATION NUMBER: 36,203
 REFERENCE/DOCKET NUMBER: 0609.3840004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 121:
 SEQUENCE CHARACTERISTICS:

US-08-340-426D-121
 Alignment Scores:
 Pred. No.: 1,05e-24 Length: 375
 Score: 303.00 Matches: 90
 Percent Similarity: 47.14% Conservative: 9
 Best Local Similarity: 42.86% Mismatches: 41
 Query Match: 6.14% Indels: 71
 DB: 2 Gaps: 6

US-09-966-880A-7 (1-2818) x US-08-340-426D-121 (1-375)
 QY 2098 CTCCTTCGCCAGCGTGGAGTGCACGCGCAGCTACTGCAACTCCACCTC 2039
 Db 5 LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArgLeu 24
 QY 2038 CTGGGTTCAAGGATTCCTCTCCCTCAGCTCCCGAGTAGCTGGGATTACAGTGCCTGC 1979

; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-450-673C-121

Alignment Scores:

Pred. No.: 1.05e-24 Length: 375
Score: 303.00 Matches: 90
Percent Similarity: 47.14% Conservative: 9
Best Local Similarity: 42.8% Mismatches: 41
Query Match: 6.14% Indels: 71
DB: 2 Gaps: 6

US-09-966-880A-7 (1-2818) x US-08-450-673C-121 (1-375)

```
QY 2098 CTCCTGTCGCCAGCGCTGAGTGGCAAGCGCAGCATCTACGCTTACTGCAACCTCCACCTC 2039
Db 5 LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArgLeu 24
QY 2038 CTGGGTTCAAGCGATTCTCCCTCCAGCTCCCAAGTAGCTGGGATTACAGGTCCCTCC 1979
Db 25 ProGlySerSerAspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCys 44
QY 1978 TACCACGCTGGCTAATT---TTTGCATTTTTCAGTACAGACGGGGTTTTCGCATGTTGC 1922
Db 45 ThrHisAlaArgLeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGly 64
QY 1921 CAGGCTGTCTCAAACTCTGACACAGGTGATCGCGCGG----- 1881
Db 65 GlnAlaGlyLeuGluLeuProThrSerAspProSerValSerAlaSerGlnSerAla 84
QY 1881 ----- 1881
Db 85 ArgTyrArgThrGlyHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArg 104
QY 1880 -----CTCGGCGCTC 1872
Db 105 ValSerLeuMetCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeu 124
QY 1871 CCAAGTCTGGGATTACAGGCTGAGCCAGCACCACCGCGGCTCTCTCTTTCTTTT 1812
Db 125 ProLysCysTrpAspTyrArgArg--AlaAlaValProGly-----LeuPheI 140
QY 1811 TTTTCTTTTGTGAGACAGCATCTGTTATGTTGCCA----- 1771
Db 140 LeuPhePheLeuArgHisArg-----CysProThrLeuThrGlnAspGluV 156
QY 1770 --GGCTGG-----CCTTGAACACCTGTGTTCAAGCCATCCTCCC 1734
Db 156 alGlnTrpCysAspHisSerSerLeuGlnProSerThrProGluIleLys-HisProPro 175
QY 1733 TCCTCAACTCCCAAGCAGCTAGACTATGACGCTGTCGCAGGTTCCGACGCTTATCAT 1674
Db 176 AlaSerAlaSerGlnValAlaGlyThrLysAspMetHisHisTyrThrTriPheLeu-IlePh 195
QY 1673 TCTGTTTGTGCACAACTCTCTT 1652
Db 195 eilePheIlePheAsnPheLeu 202
```

RESULT 7

PCT-US95-17111A-121

; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington

; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-17111A-121

Alignment Scores:
Pred. No.: 8.4e-24 Length: 397
Score: 295.00 Matches: 69
Percent Similarity: 66.67% Conservative: 11
Best Local Similarity: 57.50% Mismatches: 34
Query Match: 5.98% Indels: 6
DB: 5 Gaps: 1

US-09-966-880A-7 (1-2818) x PCT-US95-17111A-121 (1-397)

```
QY 2181 TCGTTCCTCC-----ATCCTTCTCTCCAAATATGTCTCTTCTCTCTCTCTCT 2134
Db 277 CysLeuProArgProProLysValLeuGlyLeuGlnAspValThrProThrAlaArgPro 296
QY 2133 CTTCTTTTCTTTTCTTCGACAGAGCTCTGCTCTGTCGCCCGCCAGCTCGAGTGCA 2074
Db 297 IlePheAsnPheCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGI 316
QY 2073 ACGGCAGCATCTCAGCTTACTGCAAGCTCCAGCTCTCTGGGTTCACAGCATTCCTCC 2014
Db 316 nTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLe 336
QY 2013 CAGCTCCCAAGTAGCTGGGATTACAGTGCCTGCTTACACAGCTCGCTTAATTTTGCAT 1954
Db 336 uSerLeuProSerSerTrpAspTyr-GlyHisLeuHisHisThrProLeuIlePheValP 356
QY 1953 TTTGAGTACAGACGGGTTTTCCTGCTGGCAGGCTGCTCTCAACCTCCACACACAG 1894
Db 356 heSerLeuGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSerg 376
QY 1893 GTGATCCCGCGCTCGGCTCCCAAGTGCTGGGATTACAGCGCTGAGCCACAC 1838
Db 376 lyAspProProAlaSerAlaPheGlnSerAlaGlyIleThrGlyValThrProHis 394
```

RESULT 8

US-09-058-489-22
; Sequence 22, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

; TITLE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:
Pred. No.: 1,14e-22 Length: 1079
Score: 287.00 Matches: 76
Percent Similarity: 61.54% Conservative: 12
Best Local Similarity: 53.15% Mismatches: 36
Query Match: 5.81% Indels: 19
DB: 3 Gaps: 5

US-09-966-880A-7 (1-2818) x US-09-058-489-22 (1-1079)

QY 2252 CCMAAGAGACAAATAGGATCCCTTATTGGCTCCTTACATTTTGTGGATAAGCACAA 2193
Db 954 ProAlaAspGluAsnTrpAspProThrGlyThr-----LysLysIle 967
QY 2192 TTTCCCTGC-----AATGCTTCCCATCTCTCTCCCAATATTTGTTCTCTCTCTC 2139
Db 968 TrpArgCysGluSerAsnArgSerHisThrThrIleAlaLysTyrAlaGlnTyrGlnAla 987
QY 2138 TCTCTCTCTTTTCTTTTCTTGAGACAGAGTCTTGTCTCTGCCAGGCTGGA 2079
Db 988 Ser---SerPhe-----GlnGluSerLeu-----ArgAlaGly 997
QY 2078 GTGCAACGGCAGCATCTAGCTACTGCAACCTCCACCTCCTGGGTCAAGCGATTCTCC 2019
Db 998 MetGlnTrpCysAspLeuSerLeuGlnProProProGlyPheLysArgPheSer 1017
QY 2018 TGCCCTCAGCTCCCAAGTAGCTGGGATTACAGGTGCTCTACCAGCGCTGCTATTTT 1959
Db 1018 HisLeuSerLeuProAsnSerTrpAsnTyrArgHisLeuProSerCysProThrAsnPhe 1037
QY 1958 TGCATTTTGTAGTACAGACGGGTTTGGCCATGTGGCCAGGCTGGTCTCAAACTCCTGAC 1899
Db 1038 CysIlePhe-ValGluThrGlyPheHisHisValGlyGlnAlaCysLeuGluLeuTh 1057
QY 1898 CACAGTGTATCGCCCGCGCTCGGCTCCCAAGTGTGGGATTACAGGCGTGAGCCACCA 1839
Db 1057 rSerGlyGlyLeuLeuAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisH 1077
QY 1838 CGCCCGG 1832
Db 1077 salaarg 1079

RESULT 9
US-09-288-143-168
; Sequence 168, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018p1
; CURRENT APPLICATION NUMBER: US/09/288,143
; EARLIER FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498

; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 168
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-288-143-168

Alignment Scores:
Pred. No.: 3,84e-23 Length: 99
Score: 286.00 Matches: 63
Percent Similarity: 71.88% Conservative: 6
Best Local Similarity: 65.62% Mismatches: 27
Query Match: 5.79% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-288-143-168 (1-99)
QY 2094 TGTGCGCCAGGCTGGAGTGCACGCGCACGATCTCAGCTTACTGCAACCTCCACCTCTCGG 2035
Db 2 CysCysProGlyTrpSerAlaValValArgSerTrpLeuThrAlaThrLeuAlaSerTrp 21
QY 2034 GTTCAAGCGATTCTCTCTGCC-TCAGCTCCCAAGTAGCTGGGATTACAGGTGCTCTCTAC 1976
Db 22 ValGlnAlaIleLeuMetAspSerAlaSerGlnValAlaGlyIleThrSerValHisHis 41
QY 1975 CACGCTGCTGCTAATTTTTCATTTTTCAGTACAGACGCGGTTTTCATGTTGCCAGGCT 1916
Db 42 GlnAlaGlnLeuSerPheValPheLeuValGluMetGlyLeuCysHisValGlyGlnAla 61
QY 1915 GGTCTCAAACTCTCTGACACAGGTGTCGCGCGGCTCGGCTCCCAAGTGTCTGGGATT 1856
Db 62 GlyLeuLysLeuLeuAlaSerSerAspLeuProAlaSerAlaSerGlnSerAlaGlyIle 81
QY 1855 ACAGGCGTGAGCCACACGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1808
Db 82 ThrGlyMetSerHisHisSerTrpProGluArgThrSerPheIlePhe 97

RESULT 10
US-09-227-357-193
; Sequence 193, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010p1
; CURRENT APPLICATION NUMBER: US/09/227,357
; EARLIER FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 193
LENGTH: 110
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (110)
OTHER INFORMATION: xaa equals stop translation
US-09-227-357-193

Alignment Scores:
Pred. No.: 3.15e-22 Length: 110
Score: 278.00 Matches: 61
Percent Similarity: 76.04% Conservative: 12
Best Local Similarity: 63.54% Mismatches: 16
Query Match: 5.63% Indels: 7
DB: 4 Gaps: 1

US-09-966-880A-7 (1-2818) x US-09-227-357-193 (1-110)

QY 2126 TTTTCTTTCAGACAGAGTCTGTGTCGCCAGCGCTGGAGTGCACCGCAC 2067
||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 21 PheValPheLeuArgLeuSerLeuLeu-ProLysLeuGluCysAsn----- 38

QY 2066 GATCTCAGCTTACTGCAACCTCCACCTCTGGTTCAAGCATTTCTCTGCTCAGCTC 2007
::: ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 39 -LeuGlySer--LeuGlnProProProArgPheGlnArgPheSerCysLeuSerLeu 57
QY 2006 CCAAGTAGCTGGGATTACAGGTGCCTGCTACACGCTGGCTAATTTTGCATTTTGA 1947
::: ||||||||| ||| ||||||||| ||||||||| ||||||||| |||||||||
Db 58 LeuAsnSerTrpAspTrpArgProProHisLeuAlaAsnPheCysValValSer 77
QY 1946 ACAGAGGGGTTTGGCCATGTTGCCAGGCTGCTCAAACTCCTGACACAGTGATCC 1887
||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 78 ArgGlyGlyValSerSerCysTrpProGlyTrpSerArgThrProAsp--LeuMetIle 97
QY 1886 GCCCGGCTGCCCTCCCAAGTGTGGATTACAGCG 1849
|| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 97 rgLeuProArgProProArgValLeuGlyLeuGlnAla 109

RESULT 11

US-08-687-895-1
; Sequence 1, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/08/687,895
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
US-08-687-895-1

Alignment Scores:
Pred. No.: 2.81e-21 Length: 222
Score: 271.00 Matches: 63
Percent Similarity: 50.00% Conservative: 25
Best Local Similarity: 35.80% Mismatches: 76
Query Match: 5.48% Indels: 12
DB: 1 Gaps: 5

US-09-966-880A-7 (1-2818) x US-08-687-895-1 (1-222)

QY 110 TTTCTTTACCAATTCAAAATGTCCGTGGCTAAGGTGGGCTGAGACTACCTGTGC 169
||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 110 PheValPheLeuArgLeuSerLeuLeu-ProLysLeuGluCysAsn----- 38

Db 52 PheLysPheGlnPheArgAsnValGluTyrSerSerGlyArgAsnLysThrPheLeuCys 71
Qy 170 TACGTAGTGAAGGCGGTGACAGTGCATCTCTTTTCACTGGACCTTGGTTACTTTCGC 229
Db 72 TyrValValGlnGluGlyLysGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91
Qy 230 AATAAGAAC--GGCTCCAGGTGGAAATGCTCTCTCCGCTACATCTCGGACTGGAC 286
Db 92 AspGluHisAlaAlaHisAlaGluGluAlaPheAsnThrIleLeuPro--Ala 110
Qy 287 CTAGACCTGGCGCTCTACCGCTGACCTGCTGCTGACCTGCTGACCTGCTGAC 346
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerProCysAlaAla 130
Qy 347 TGTGCGGACATGTGCGGCTCTTCTGCGAGGAAACCCCACTCTGAGGATCTC 406
Db 131 CysAlaAspArgIle***LysThrLeuSerLysThrLysAsnLeuArgLeuLeu 150
Qy 407 ACCGCGGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAG-----GGGCTG 457
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluLeuGlnAlaAlaLeu 166
Qy 458 CGCGGCTGACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTAC 517
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186
Qy 518 TGTGGAATACTTTGTAGAAAACCATGAA-----AGAACTTTCAAGCTGGGAA 568
Db 187 ValTrpGlnAsnPheValGlnGluGluGluGluGluGluGluGluGluGluGlu 206
Qy 569 GGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATC 616
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIle 222

RESULT 12

US-09-040-482-1

; Sequence 1, Application US/09040482
; Patent No. 5916556
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,482
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/687,895
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
; US-09-040-482-1

Alignment Scores:
Pred. No.: 2,81e-21 Length: 222
Score: 271.00 Matches: 63
Percent Similarity: 50.00% Conservative: 25
Best Local Similarity: 35.80% Mismatches: 76
Query Match: 5.48% Indels: 12
DB: 2 Gaps: 5

US-09-966-880A-7 (1-2818) x US-09-040-482-1 (1-222)

Qy 110 TTTCTTTTACCAATTCAAAATGTCCGCTGGGCTAAGGTCGGCTGAGACCTACCTGTGC 169
Db 52 PheLysPheGlnPheArgAsnValGluTyrSerSerGlyArgAsnLysThrPheLeuCys 71
Qy 170 TACGTAGTGAAGGCGGTGACAGTGCATCTCTTTTCACTGGACCTTGGTTACTTTCGC 229
Db 72 TyrValValGlnGluGlyLysGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91
Qy 230 AATAAGAAC--GGCTCCAGGTGGAAATGCTCTCTCCGCTACATCTCGGACTGGAC 286
Db 92 AspGluHisAlaAlaHisAlaGluGluAlaPheAsnThrIleLeuPro--Ala 110
Qy 287 CTAGACCTGGCGCTCTACCGCTGACCTGCTGCTGACCTGCTGACCTGCTGAC 346
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerProCysAlaAla 130
Qy 347 TGTGCGGACATGTGCGGCTCTTCTGCGAGGAAACCCCACTCTGAGGATCTC 406
Db 131 CysAlaAspArgIle***LysThrLeuSerLysThrLysAsnLeuArgLeuLeu 150
Qy 407 ACCGCGGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAG-----GGGCTG 457
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluLeuGlnAlaAlaLeu 166
Qy 458 CGCGGCTGACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTAC 517
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186
Qy 518 TGTGGAATACTTTGTAGAAAACCATGAA-----AGAACTTTCAAGCTGGGAA 568
Db 187 ValTrpGlnAsnPheValGlnGluGluGluGluGluGluGluGluGluGluGlu 206
Qy 569 GGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATC 616
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIle 222

RESULT 13

US-09-227-357-285
; Sequence 285, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929

EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 285
LENGTH: 122
TYPE: PRT
ORGANISM: Homo sapiens
US-09-227-357-285

Alignment Scores:
Pred. No.: 7,19e-19 Length: 122
Score: 248.00 Matches: 56
Percent Similarity: 61.86% Conservative: 4
Best Local Similarity: 57.73% Mismatches: 36
Query Match: 5.02% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-227-357-285 (1-122)
QY 2054 CTGCAACCTCCACCTCCTGGTTCAAGCGATTCTCTGCTCAGCTCCAGCTCCCAAGTAGCTGG 1995
Db :::::|||||
1 MetGlnAlaLeuProGlyPheLysGlnPheSerCysLeuSerLeuProSerArgTrr 20
QY 1994 GATTACAGGTGCTGTACACGCTGGCTAATTTTGGATTTTGGATTTGAGTACAGACGGGGTT 1935
Db |||||
21 AspTyrGlyCysAlaThrGlnHisProAlaAsnPheCysIlePheArgArgAspArgVal 40
QY 1934 TTGGCATGTTGGCCAGGCTGCTCAAACTCTCTGACACAGGTGATCCCGCGCTCGGC 1875
Db |||||
41 -SerHisValGlyGlnAlaGlyLeuLysLeuThrSerValAspProProAlaTrpAl 60
QY 1874 CTCCCAAGTGTGGGATTACAGCGTGAGCCACACCGCCGCTCTCTCTTTCTTTT 1815
Db |||||
60 aSerGlnSerAlaGlyIleThrGlyLysSerHisCysAlaGlnLeuHisCysCysPh 80
QY 1814 TTTTCTTTTGTGAGACAGCATCTTGTATGTTGCCAGGCTG 1766
Db |
80 eLeuLeuValLysArgAspGlnProLeuGluLysCysLeuArgLeu 96

RESULT 14

US-08-687-895-3
Sequence 3, Application US/08687895
Patent No. 5747319
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Hawkins, Phillip R.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,895
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0109 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 436941
US-08-687-895-3

Alignment Scores:
Pred. No.: 1,51e-18 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 4.95% Indels: 10
DB: 1 Gaps: 3

US-09-966-880A-7 (1-2818) x US-08-687-895-3 (1-116)
 QY 305 TACCGCGTCCACCTGGTTCACCTCTCGAGCCCTGCTACGAC-----TGTGCCCGACAT 358
 Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26
 QY 359 GTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTC 418
 Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
 QY 419 TACTCTCTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTCGCGCGCTGCAC 469
 Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
 QY 470 CGCGCGGGGTGCAATAGCCATCATGACCTTCAAGATATTATTTTACTGCTGGAATACT 529
 Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81
 QY 530 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCTGGGAAGGCTGCATGAAAATTCAGTT 589
 Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGln 101
 QY 590 CGTCTCTCAGACAGCTTCGGCGCATCCTT 619
 Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 15

US-08-816-241-3
 ; Sequence 3, Application US/08816241
 ; Patent No. 5804185
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Goli, Surya K.
 ; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/816,241
 ; FILING DATE: Filed Herewith
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0239 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-853-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 116 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 436941
 ; US-08-816-241-3

Alignment Scores:
 Pred. No.: 1-51e-18 Length: 116
 Score: 245.00 Matches: 51
 Percent Similarity: 61.82% Conservative: 17
 Best Local Similarity: 46.36% Mismatches: 32
 Query Match: 4.95% Indels: 10
 DB: 1 Gaps: 3
 US-09-966-880A-7 (1-2818) x US-08-816-241-3 (1-116)
 QY 305 TACCGCGTCCACCTGGTTCACCTCTCGAGCCCTGCTACGAC-----TGTGCCCGACAT 358
 Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26
 QY 359 GTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTC 418
 Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
 QY 419 TACTCTCTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTCGCGCGCTGCAC 469
 Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
 QY 470 CGCGCGGGGTGCAATAGCCATCATGACCTTCAAGATATTATTTTACTGCTGGAATACT 529
 Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81
 QY 530 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCTGGGAAGGCTGCATGAAAATTCAGTT 589
 Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGln 101
 QY 590 CGTCTCTCAGACAGCTTCGGCGCATCCTT 619
 Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

Search completed: June 14, 2003, 18:59:19
 Job time : 37.1417 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:26:00 ; Search time 90.7352 Seconds

(without alignments)
6641.044 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 4948

Sequence: 1 agagaaccatcatattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Fgapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp

-Q/cgn2_1/USPTO_spool/US09966880/runat_14062003_175526_10413/app.query.fasta_1.9493

-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100

-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09966880 @CGN_1_1_105.@runat_14062003_175526_10413

-NCPU=6 -ICPU=3 -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*

7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap.*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*

9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pap.*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1086	21.9	198	US-09-966-880A-8
2	1008	20.4	198	US-09-966-880A-2
3	390	7.9	384	US-09-729-674-174
4	381.5	7.7	239	US-09-800-729-193

381.5	7.7	310	10	US-09-800-729-190	Sequence 190, App
369.5	7.5	222	10	US-09-925-300-1639	Sequence 1639, App
365.5	7.4	109	9	US-09-989-919-105	Sequence 105, App
365	7.4	94	9	US-10-001-835-135	Sequence 135, App
361	7.3	107	9	US-09-989-920-233	Sequence 233, App
359	7.3	144	9	US-10-011-585A-145	Sequence 145, App
358	7.3	100	9	US-10-016-157A-171	Sequence 171, App
351.5	7.1	126	9	US-10-082-830-207	Sequence 207, App
351	7.1	172	10	US-09-764-864-1058	Sequence 1058, App
349.5	7.1	121	9	US-09-989-920-198	Sequence 198, App
343	6.9	198	9	US-09-986-480-433	Sequence 433, App
338.5	6.9	123	9	US-09-764-868-888	Sequence 888, App
338.5	6.9	170	9	US-09-986-480-282	Sequence 282, App
332	6.7	115	10	US-09-995-494-79	Sequence 79, App
332	6.7	132	9	US-10-012-896-573	Sequence 573, App
332	6.7	132	9	US-09-895-793-573	Sequence 573, App
332	6.7	132	9	US-09-895-814-573	Sequence 573, App
332	6.7	132	10	US-09-759-143-573	Sequence 573, App
332	6.7	132	10	US-09-780-669-573	Sequence 573, App
332	6.7	132	10	US-09-822-827-573	Sequence 573, App
332	6.7	135	9	US-10-012-896-884	Sequence 884, App
332	6.7	135	9	US-09-895-793-884	Sequence 884, App
332	6.7	135	9	US-09-895-814-884	Sequence 884, App
332	6.7	135	10	US-09-759-143-884	Sequence 884, App
332	6.7	135	10	US-09-780-669-884	Sequence 884, App
332	6.7	135	10	US-09-822-827-884	Sequence 884, App
328.5	6.7	122	9	US-10-074-475-190	Sequence 190, App
320.5	6.5	118	10	US-09-764-898-157	Sequence 157, App
319.5	6.5	117	9	US-10-001-835-203	Sequence 203, App
315.5	6.4	121	10	US-09-764-877-1641	Sequence 1641, App
312.5	6.3	109	9	US-10-001-835-154	Sequence 154, App
311	6.3	90	9	US-10-001-873-34	Sequence 34, App
310	6.3	93	9	US-09-764-891-5337	Sequence 5337, App
309	6.3	86	9	US-10-011-585A-154	Sequence 154, App
309	6.3	361	10	US-09-995-494-107	Sequence 107, App
308	6.2	157	9	US-10-001-835-197	Sequence 197, App
304.5	6.2	100	9	US-10-016-157A-187	Sequence 187, App
303	6.1	375	9	US-10-146-130-2	Sequence 2, App
303	6.1	375	9	US-09-964-667-2	Sequence 2, App
303	6.1	375	9	US-09-872-968-2	Sequence 2, App
303	6.1	375	9	US-10-092-934-10	Sequence 10, App

ALIGNMENTS

RESULT 1
US-09-966-880A-8
; Sequence 8, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-880A-8

Alignment Scores:

Pred. No.: 1.43e-102 Length: 198
 Score: 1086.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 21.95% Indels: 0
 DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-966-880A-8 (1-198)

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 139
 Db 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
 QY 140 GCTAAGGTCGGCTGAGACCTACCTGCTAGTAGTGAAGAGGCGTGCACAGTGTCTACA 199
 Db 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 QY 200 TCCTTTTCACTGACCTTTCGTTATCTTCGCATATAGAACGGTGCACAGTGGAAATGCTC 259
 Db 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTACCTGG 319
 Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 QY 320 TTCACCTCTCGAGCCCTCTGCTACGACTGTGCGCCGCTGCTACCGGCTACCTGG 379
 Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 QY 380 ACCCCCAACCTCAGTCTGAGGATCTTACCGCGGCTCTACTTCTGTGAGGACCGCAAG 439
 Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 QY 440 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 499
 Db 121 AlaGluProGluGlyLeuArgLysHisArgAlaGlyValGlnIleAlaIleMetThr 140
 QY 500 TTCAAAGATTATTTTACTGCTGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 559
 Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCCT 619
 Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 QY 620 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCTGTACTTTGGGACTT 673
 Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2

US-09-966-880A-2
 ; Sequence 2, Application US/09966880A
 ; Patent No. US20020164743A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Honjo, Tasuku
 ; TITLE OF INVENTION: Muramatsu, Masamichi
 ; FILE REFERENCE: NOVEL CYTIDINE DEAMINASE
 ; CURRENT APPLICATION NUMBER: US/09/966,880A
 ; CURRENT FILING DATE: 2001-09-28
 ; PRIOR APPLICATION NUMBER: PCT/JP00/01918
 ; PRIOR FILING DATE: 2000-03-28
 ; PRIOR APPLICATION NUMBER: JP 11-371382
 ; PRIOR FILING DATE: 1999-12-27
 ; PRIOR APPLICATION NUMBER: JP 11-178999
 ; PRIOR FILING DATE: 1999-06-24
 ; PRIOR APPLICATION NUMBER: JP 11-87192
 ; PRIOR FILING DATE: 1999-03-29
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 198
 ; TYPE: PRT

ORGANISM: Mus musculus

US-09-966-880A-2

Alignment Scores:
 Pred. No.: 1.51e-94 Length: 198
 Score: 1008.00 Matches: 183
 Percent Similarity: 95.94% Conservative: 6
 Best Local Similarity: 92.89% Mismatches: 8
 Query Match: 20.37% Indels: 0
 DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-966-880A-2 (1-198)

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 139
 Db 1 MetAspSerLeuLeuMetLysGlnLysLysPheLeuTyrHisPheLysAsnValArgTrp 20
 QY 140 GCTAAGGTCGGCTGAGACCTACCTGCTAGTAGTGAAGAGGCGTGCACAGTGTCTACA 199
 Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 QY 200 TCCTTTTCACTGACCTTTCGTTATCTTCCCAATAAGAACGGTGCACAGTGGAAATGCTC 259
 Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeuLeu 60
 QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTACCTGG 319
 Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 QY 320 TTCACCTCTCGAGCCCTCTGCTACGACTGTGCGCCGACATGTGCGGACTTCTGCGAGGG 379
 Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
 QY 380 ACCCCCAACCTCAGTCTGAGGATCTTACCGCGGCTCTACTTCTGTGAGGACCGCAAG 439
 Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 QY 440 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGGCTCTACTTGTAGAAAACCATGAAGAACTTTCAAA 559
 Db 121 AlaGluProGluGlyLeuArgLysHisArgAlaGlyValGlnIleGlyIleMetThr 140
 QY 500 TTCAAAGATTATTTTACTGCTGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 559
 Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160
 QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCCT 619
 Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180
 QY 620 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCTGTACTTTGGGA 670
 Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3

US-09-729-674-174
 ; Sequence 174, Application US/09729674
 ; Patent No. US2001003935A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: Lavallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Steininger II, Robert J.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fechtel, Kim
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

QY		170	TACGTAGTGAAG-----AGCGCTGACAGTGCTATACCTCTTTTCACCTGGACTTTGGTTAT	223
Db		:	::: :::	
QY		69	PheThrValGluglyLeuLysArgSerValValSerTrpLysThr	86
Db		:	::: :::	
QY		224	CATTGCCAAT-----AGAAGCGCTGCACCTGGAAATGGCTCTTCCCTCCGCTAC	271
Db		:	::: :::	
QY		87	PheArgAsnGluValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp	106
Db		:	::: :::	
QY		272	ATCTCGGACTGGGACCTAGACCTGGCGCGCTGCTACCGGCTCACCTGGTTCACTCCCTGG	331
Db		:	::: :::	
QY		107	PheCysAspAspIleLeuSerProAsnThrLysTyrglnValTrpTrpTrpSerTrp	126
Db		:	::: :::	
QY		332	AGCCCCCTCTACGACTGTGCCGCACATGTGGCGCGACTTTCTCGAGGGAACCCCNAACCTC	391
Db		:	::: :::	
QY		127	SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal	146
Db		:	::: :::	
QY		392	AGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCGTAGCCCGAG	451
Db		:	::: :::	
QY		147	AsnLeuThrIlePheThrAlaArgLeuTyrrTyrrPheGln---TyrProCysTyrglnGlu	165
Db		:	::: :::	
QY		452	GGCTGCGCGCGCTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTCAAAGATTAT	511
Db		:	::: :::	
QY		166	GlyLeuArgSerLeuSerGlnGlyValAlaValGluIleMetAspTyrgluAspPhe	185
Db		:	::: :::	
QY		512	TTTTACTGCTGGNACTTTTCTAGAAAACCATCAAGAACTTTTCAAAGCTCGGAAGGG	571
Db		:	::: :::	
QY		186	LysTyrcysTrpGluAsnPheValTyraasnAspAsnGluProPheLysProTrpLysGly	205
Db		:	::: :::	
QY		572	CTGCATGAAAAATTCAGTTCGCTCCACACAGCTTCGGCGCATCCCT	619
Db		:	::: :::	
QY		206	LeuLysThrAsnPheArgLeuLeuLysArgLeuArgGluSerLeu	22
Db		:	::: :::	

Db 20 rtleLeuAlaTyHisAsnLeuCysLeuLeuGlySerSerAsnSerProAlaSerGlySe 40
QY 2006 CCAAGTAGCTGGATTACAGGTGCTGCTACACCGCTGGCTAAATTTTGCATTGTGAGT 1947
Db 40 rGlnValAlaGlyIleThrGlyMetCysHisThrArgLeuIlePheValPheLeuVa 60
QY 1946 ACAGACGGGGTTTGGCATGTGTGGCAGGCTGCTCAAACTCCTGACACAGGTGATCC 1887
Db 60 lGluThrGlyTyLeuHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspPr 80
QY 1886 GCCCGCTGGCTGCCAAGTGTGGGATTACAGCGGTGAGCCAGCCAGCCGGGCC-- 1829
Db 80 oProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerArgHisAlaTrpProSe 100
QY 1828 -TCTCTCTTCTTTTTTTTTT 1808
Db 100 rSerAlaPheIleHisIlePhe 107

RESULT 8

US-10-001-835-135
; Sequence 135, Application US/10001835
; Patent No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 135
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-135

Alignment Scores:
Pred. No.: 1,36e-28 Length: 94
Score: 365.00 Matches: 71
Percent Similarity: 84.78% Conservative: 7
Best Local Similarity: 77.17% Mismatches: 14
Query Match: 7.39% Indels: 0
DB: Gaps: 0

US-09-966-880A-7 (1-2818) x US-10-001-835-135 (1-94)

QY 2101 TTGCTCTTTCGCCAGGTGGAGTGCACCGCAGCATCTCAGCTTACTGCAACCTCCAC 2042
Db 3 LeuThrLeuSerProArgLeuGluCysSerGlyThrValSerAlaHisCysAsnLeuHis 22
QY 2041 CTCCTGGGTCAAGCATTTCTCTGCTCAGCCTCCCAAGTAGCTGGGATTACAGTGCC 1982
Db 23 LeuLeuGlySerSerAspSerProAlaSerValSerAlaValAlaGlyThrThrGlyVal 42
QY 1981 TGCTACACGCTGGCTAATTTTGCATTTTGTAGTACACACGGGGTTTGGCATGTGGC 1922
Db 43 ArgHisHisAlaTrpLeuIlePheIlePheLeuValGluThrValPheCysHisValGly 62
QY 1921 CAGGCTGGTCTCAAACTCCTGACACAGGTGATCCCGCGGTGCTGGCCTCCCAAAAGTCT 1862
Db 63 GlnAlaGlyLeuLysLeuLeuThrSerGlyAspProProThrSerAlaSerAlaSerThr 82
QY 1861 GGGATTACAGCGGTGAGCCACACGCGCCGCTCT 1826
Db 83 GlyIleThrGlyMetSerHisCysAlaTrpProSer 94

RESULT 9

US-09-989-920-233
; Sequence 233, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 233
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-233

Alignment Scores:
Pred. No.: 3.7e-28 Length: 107
Score: 361.00 Matches: 76
Percent Similarity: 79.25% Conservative: 18
Best Local Similarity: 71.70% Mismatches: 19
Query Match: 7.31% Indels: 3
DB: Gaps: 1

US-09-966-880A-7 (1-2818) x US-09-989-920-233 (1-107)

QY 2129 TTTTNTTTTTTTTTTCTGAGA-----CAGAGTCTTCTCTTCTGCTGCCAGGTGGAGTG 2076
Db 1 PhePhePhePheLeuLeuLeuPheCysAspSerLeuAla-LeuSerProArgLeuGlnCy 20
QY 2075 CAACGGCAGATCTCAGCTTACTGCAACCTCCACCTCCCTGGGTTCAGCGGATTCCTCTGC 2016
Db 20 sSerGlyThrIleSerAlaHisCysAsnLeuLeuGlySerSerAsnSerProva 40
QY 2015 CTCAGCTCCCAAGTAGCTGGATTACAGTGTGCTTACACGCTGGCTAATTTTTCG 1956
Db 40 lSerAlaSerTrpValAlaGlyThrGlyAlaCysHisHisAlaTrpLeuThrPheVa 60
QY 1955 ATTTGAGTACAGACGGGGTTTGGCATGTGTGCCAGGTGGTCTCAAACTCCTGACCAC 1896
Db 60 lPheLeuValGluThrGlyPheHisHisValGlyGlnAlaGlyLeuGluPheLeuThrSe 80
QY 1895 AGGTGATCCGCGCGGCTCGGCTCCCAAAAGTGTGGGATTACAGGCTGAGCCAGCCACGC 1836
Db 80 rGlyAspProProAlaLeuAlaSerGlnSerAlaGluIleThrGlyValSerHisArgAl 100
QY 1835 CCGGCCCTCTCTTT 1820
Db 100 atrpProValCysPhe 105

RESULT 10

US-10-011-585A-145
; Sequence 145, Application US/10011585A
; Publication No. US20030039986A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
; FILE REFERENCE: DEX-0261
; CURRENT APPLICATION NUMBER: US/10/011,585A
; CURRENT FILING DATE: 2002-03-14

Db 29 LeuAlaMetSerGlnSer--IleMetAlaLeuSerProArgLeuGluCysAsnGlyAlaVal 48
QY 2064 TCTCAGCTTACTGCAACCTCCACCTCCTGGGTTCAAGGATTCCTCTGCTCAGGCTCCC 2005
:||||| :||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 48 alserGlyHisCysAsnProCysLeuProGlySerSerAspSerProProSerAlaSerG 68
QY 2004 AGTAGCTGGGATTACAGTGCTGCTACACGCCCTGGCTAAATTTTGGATTGTAGTAC 1945
:||||| :||||| ||| :||||| ||||||| ||||||| ||||||| ||| :|||
Db 68 InValAlaGlyIleThrGlyThrCysHisAlaArgLeuIlePheValPheLeuValG 88
QY 1944 AGACGGGGTTTGCCCATGTGGCAGGCTGGCTCAAACTCCTGACCACAGGTGATCCGC 1885
: : ||||| ||||||| ||||||| ||||||| :||||| :||||| :||||| :|||
Db 88 luMetGlyPheHisValGlyGlnAlaGlyLeuGluLeuThrSerGlyAspLeuP 108
QY 1884 CCGGCTCGGCCCTCCCAAGTGTGGATTACAGGCTGAGCCACACCGCCGCCGCC 1829
|| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 108 roThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgPro 126

RESULT 13
US-09-764-864-1058
; Sequence 1058, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1058
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1058

Alignment Scores:
Pred. No.: 4,8e-27 Length: 172
Score: 351.00 Matches: 69
Percent Similarity: 80.00% Conservativeness: 7
Best Local Similarity: 72.63% Mismatches: 19
Query Match: 7.11% Indels: 1
Gaps: 0
US-09-966-880A-7 (1-2818) x US-09-764-864-1058 (1-172)

QY 2110 GACAGAGTCTTGCTCTTGCGCCAGGCTGGAGTGCAACGGCAGCATCTCAGCTTACTGC 2051
||| ||| ||||||| ||||||| ||||||| ||||||| :||| :||| :||| :||| :||| :|||
Db 1 AspGluValSerLeuLeuSerProArgLeuValCysAsnGlyMetMetSerAlaHisCys 20
QY 2050 AACCTCCACCTCTCTGGGTTCAAGCGATTCTCCTGCTCAGCCTCCCAAGTAGCTGGGATT 1991
||| ||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 21 LysLeuArgLeuProGlySerSerAspSerProThrSerAlaSerGlnValAlaGlyIle 40
QY 1990 ACAGTGCTGTACACGCCCTGGCTAAATTTTGGCATTTTGGATACAGAGCGGGTTTTCG 1931
||||| ||||| :||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 41 ThrGlyThrCysHisAlaTrpLeuIlePheValLeuLeuValGluMetGlyTyrAsp 60
QY 1930 CATGTTGGCAGGCTGGCTCAAACTCCTGACCACAGGTATCCGCCCGG-CTCGGCCCTC 1872
||||| :||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 HisLeuGlyGlnAlaAlaLeuLysLeuLeuThrSerGlyAspProProThrSerGlyLeu 80
QY 1871 CCAAACTGCTGGATTACAGCGCTGAGCCACACCGCCGCCCTC 1827
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 81 ProLysCysTrpAspTrpArgHisGluProProLeuProAlaLeu 95

RESULT 14
US-09-989-920-198
; Sequence 198, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto

; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 198
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-198
Alignment Scores:
Pred. No.: 5,92e-27 Length: 121
Score: 349.50 Matches: 78
Percent Similarity: 68.46% Conservativeness: 11
Best Local Similarity: 60.00% Mismatches: 32
Query Match: 7.08% Indels: 9
Gaps: 3
US-09-966-880A-7 (1-2818) x US-09-989-920-198 (1-121)
QY 2119 TTTTCTGACAGAGCTCTGCTTGTGCGCCAGGCTGGAGTGCAACGGCAGCATCTCA 2060
||||| ||| ||| ||||||| ||||||| ||||||| :||| :||| :||| :||| :||| :|||
Db 1 PhePheAlaAspGluValSerArgLeuSerProGlyLeuGluCysSerGlyValIleSer 20
QY 2059 GCTTACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCCTCAGCTCCAGCTCCCAAGTA 2000
||||| :||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 21 AlaHisCysAsnPheHisLeuLeuGlySerSerSerSerProAlaSerAlaSerGlnVal 40
QY 1999 GCTGGATTACAGGTGCTGCTACACAGCTGGCTAAATTTTGGCATTTTGGATACAGACG 1940
||||| ||||||| ||||||| :||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 41 AlaGluIleThrGlyAlaCysHisProThrTrpLeuIlePheValIleLeuValGluThr 60
QY 1939 GGGTTTGGCATTTGGCCAGGCTGCTCAAACTCCTGACCACAGGTATCCGCCCGG 1880
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 GlyPheHisValGlyGlnAlaAsp---AlaLeuLeuThrSerGlyAspProProPhe 79
QY 1879 TCGGCTCCCAAAGTCTGGATTACAGCGGTACAGCGCACACCGCCGCCCTCTCTCTT 1820
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 80 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgProAla----- 97
QY 1819 CTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1760
||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 98 -----AsnThrPheAlaLeuThrThrLeuGlyLeuLeuTyrlLysIleValMet 113
QY 1759 AACACCTGTGTTCAACCCATCTCTCTCT 1730
:||||| |||||
Db 114 -----IleAlaMetGluValLeuProPro 121
RESULT 15
US-09-986-480-433
; Sequence 433, Application US/09986480
; Publication No. US20030027999A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/12788
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 433

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 18, 2003, 23:33:05 ; Search time 7195.7 Seconds
(without alignments)
11397.322 Million cell updates/sec
Title: US-09-966-880A-7
Perfect score: 2818
Sequence: 1 adagaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 205460 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sv:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2791	99.0	2791	9	AB040431	AB040431 Homo sapi
2	2174.6	77.2	11204	9	AB040430	AB040430 Homo sapi
3	2174.6	77.2	71132	9	AC092184	AC092184 Homo sapi
4	1820	64.6	1837	9	BC006296	BC006296 Homo sapi
5	603.4	21.4	2440	10	AF132979	AF132979 Mus muscu
6	595.4	21.1	597	12	AF529828	AF529828 Mus muscu
7	594.4	21.1	596	9	AF529815	AF529815 Homo sapi
8	594.4	21.1	596	9	AF529816	AF529816 Homo sapi
9	594.4	21.1	596	9	AF529819	AF529819 Homo sapi
10	594.4	21.1	596	9	AF529820	AF529820 Homo sapi
11	594.4	21.1	596	9	AF529821	AF529821 Homo sapi
12	594.4	21.1	596	9	AF529822	AF529822 Homo sapi
13	594.4	21.1	596	9	AF529823	AF529823 Homo sapi
14	594.4	21.1	596	9	AF529824	AF529824 Homo sapi
15	594.4	21.1	596	9	AF529825	AF529825 Homo sapi
16	594.4	21.1	596	9	AF529826	AF529826 Homo sapi
17	594.4	21.1	596	9	AF529827	AF529827 Homo sapi
18	594.4	21.1	596	12	AF529829	AF529829 Mus muscu
19	594.4	21.1	596	12	AF529830	AF529830 Mus muscu
20	594.4	21.1	596	12	AF529831	AF529831 Mus muscu
21	594.4	21.1	596	12	AF529833	AF529833 Mus muscu
22	594.4	21.1	596	12	AF529835	AF529835 Mus muscu
23	594.4	21.1	596	12	AF529837	AF529837 Mus muscu
24	594.4	21.1	596	12	AF529839	AF529839 Mus muscu
25	593.4	21.1	595	12	AF529834	AF529834 Mus muscu
26	592.8	21.0	596	9	AF529817	AF529817 Homo sapi
27	592.8	21.0	596	9	AF529818	AF529818 Homo sapi
28	592.8	21.0	596	12	AF529832	AF529832 Mus muscu
29	592.8	21.0	596	12	AF529836	AF529836 Mus muscu
30	592.8	21.0	596	12	AF529838	AF529838 Mus muscu
31	592.2	21.0	597	12	AF529841	AF529841 Crictetu
32	592.2	21.0	597	12	AF529842	AF529842 Crictetu
33	592.2	21.0	597	12	AF529843	AF529843 Crictetu
34	592.2	21.0	597	12	AF529844	AF529844 Crictetu
35	592.2	21.0	597	12	AF529845	AF529845 Crictetu
36	592.2	21.0	597	12	AF529846	AF529846 Crictetu
37	592.2	21.0	597	12	AF529847	AF529847 Crictetu
38	592.2	21.0	597	12	AF529848	AF529848 Crictetu
39	592.2	21.0	597	12	AF529849	AF529849 Crictetu
40	592.2	21.0	597	12	AF529850	AF529850 Crictetu
41	592.2	21.0	597	12	AF529851	AF529851 Crictetu
42	592.2	21.0	597	12	AF529852	AF529852 Crictetu
43	592.2	21.0	597	12	AF529853	AF529853 Crictetu
44	592.2	21.0	597	12	AF529854	AF529854 Crictetu
45	586.2	20.8	591	12	AF529855	AF529855 Crictetu

ALIGNMENTS

RESULT 1
AB040431
LOCUS
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,
complete CDS.
ACCESSION AB040431
VERSION AB040431.1 GI:9988409
KEYWORDS AID; activation-induced cytidine deaminase; Human AID.
SOURCE Homo sapiens CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE Isolation, tissue distribution, and chromosomal localization of the

JOURNAL MEDLINE REFERENCE AUTHORS	human activation-induced cytidine deaminase (AID) gene Genomics 68 (1), 85-88 (2000)	Db
TITLE	2 (sites)	Qy
JOURNAL MEDLINE REFERENCE AUTHORS	Revy, P., Muto, T., Levv, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lageouse, K., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A.	Db
TITLE	Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)	Qy
JOURNAL MEDLINE REFERENCE	Cell 102 (5), 565-575 (2000)	Db
TITLE	20460541	Qy
JOURNAL MEDLINE REFERENCE	3 (bases 1 to 2791)	Db
TITLE	Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.	Qy
JOURNAL	Submitted (18-NAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp, Tel: 81-75-753-4371 (ex. 4371), Fax: 81-75-753-4388)	Db
FEATURES	Location/Qualifiers	Db
source	1. .2791	Qy
gene	/organism="Homo sapiens"	Db
CDS	/db_xref="taxon:9606"	Db
	1. .2791	Qy
	/gene="AID"	Db
	77. .673	Qy
	/gene="AID"	Db
	/codon_start=1	Qy
	/product="activation-induced cytidine deaminase"	Db
	/protein_id="BAB12721.1"	Qy
	/db_xref="GI:9988410"	Db
	/translation="MDSLNNRRKLYQFNVRWAKRRRETYLCYVVKRRDSATFSLDFGLRNKNGCHVELLFRLYISDMDLPGRCYRVWFTSKSPCYDCARHVPDLRNP NLSLRIFTARLPFCEDEKRAPEGLRLHRAGVQIAIMTKDYFYCNWTFVFNHERFTK AWGLHNSVRSLRQRLPLPFLREVDLUDAPRTLGL"	Qy
BASE COUNT	842 a 548 c 625 g 776 t	Db
ORIGIN		Qy
Query Match	99.0% Score 2791; DB 9; Length 2791;	Db
Best Local Similarity	100.0% Pred. No. 0;	Qy
Matches 2791; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	Db
Qy	4 GAACATCATTAATGAAGTGAGATTTTCTGCGCTGAGACTTGCAGGGAGCAAGA 63	Qy
Db		Db
Qy	1 GAACATCATTAATGAAGTGAGATTTTCTGCGCTGAGACTTGCAGGGAGCAAGA 60	Qy
Db		Db
Qy	64 CACTCTGGACACCACTATGAGAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATT 123	Qy
Db		Db
Qy	61 CACTCTGGACACCACTATGAGAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATT 120	Qy
Db		Db
Qy	124 CAAAAATGTCGCTGGCTAAGGTCGCGTGAGACCTACCTGCTGCTAGCTAGAGAG 183	Qy
Db		Db
Qy	121 CAAAATGTCGCTGGCTAAGGTCGCGTGAGACCTACCTGCTGCTAGCTAGAGAG 180	Qy
Db		Db
Qy	184 GCCTGACAGTGCATACCTCTTTTCACTGGACTTTGGTTAFCITTCGCAATAAGAACGCGTG 243	Qy
Db		Db
Qy	181 GCCTGACAGTGCATACCTCTTTTCACTGGACTTTGGTTAFCITTCGCAATAAGAACGCGTG 240	Qy
Db		Db
Qy	244 CCAGCTGGGAATTCCTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGTG 303	Qy
Db		Db
Qy	241 CCAGCTGGGAATTCCTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGTG 300	Qy
Db		Db
Qy	304 CTACCGCGTCACTGGTTCACTTCCTGGAGCCCTGCTAGCACTGTGCCGACATGTGGC 363	Qy
Db		Db
Qy	301 CTACCGCGTCACTGGTTCACTTCCTGGAGCCCTGCTAGCACTGTGCCGACATGTGGC 360	Qy
Db		Db
Qy	364 CGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTT 423	Qy
Db		Db
Qy	361 CGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTT 420	Qy
Db		Db
Qy	424 CTCTGAGGACCGCAAGCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCA 483	Qy
Db		Db

Db	421 CTCTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGCGGCTGCAACCGCGGGGTGCA 480	Qy
Qy	484 AATAGCCATCATGACCTTCAAGATATATTTTACTGCTGAATACTTTTCTAGAAAACCA 543	Db
Db	481 AATAGCCATCATGACCTTCAAGATATATTTTACTGCTGAATACTTTTCTAGAAAACCA 540	Qy
Qy	544 TGAAGAACCTTTCAAGGCTTGGGAAGGCTGCATGAAAATTCAGTTGCTCTCTCCAGACA 603	Db
Db	541 TGAAGAACCTTTCAAGGCTTGGGAAGGCTGCATGAAAATTCAGTTGCTCTCTCCAGACA 600	Qy
Qy	604 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACCGCATTTCTGAC 663	Db
Db	601 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACCGCATTTCTGAC 660	Qy
Qy	664 TTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACCATGAAATATCTCTGCTCAAG 723	Db
Db	661 TTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACCATGAAATATCTCTGCTCAAG 720	Qy
Qy	724 ACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCACTTT 783	Db
Db	721 ACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCACTTT 780	Qy
Qy	784 CTTAGAGTTTACAGAAAATATTTATATAGGACTCTTTTAAAAGATCTATGCTTGA 843	Db
Db	781 CTTAGAGTTTACAGAAAATATTTATATAGGACTCTTTTAAAAGATCTATGCTTGA 840	Qy
Qy	844 ATAGAGAAGAACACAGGCTGCGCAGGAGGCTGCTCAATTTGCTGAGTTTGAATGCA 903	Db
Db	841 ATAGAGAAGAACACAGGCTGCGCAGGAGGCTGCTCAATTTGCTGAGTTTGAATGCA 900	Qy
Qy	904 ACATTGTCCTTACTGCGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGCAA 963	Db
Qy	901 ACATTGTCCTTACTGCGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGCAA 960	Qy
Qy	964 CGTTTTCTATGACTTTTAGTGGATGAGAGCAGAGGTAGATCTTAAAAGCATGGTG 1023	Db
Db	961 CGTTTTCTATGACTTTTAGTGGATGAGAGCAGAGGTAGATCTTAAAAGCATGGTG 1020	Qy
Qy	1024 AGAGGATCAAAATGTTTTATATCAACATCTTTTATTTGATTGATTTCAGTTAAACAGT 1083	Db
Db	1021 AGAGGATCAAAATGTTTTATATCAACATCTTTTATTTGATTGATTTCAGTTAAACAGT 1080	Qy
Qy	1084 GGTGTTAGTGATAGATTTTCTATTTCTTTCCCTTGACGTTTACTTTTCAAGTAACACAA 1143	Db
Db	1081 GGTGTTAGTGATAGATTTTCTATTTCTTTCCCTTGACGTTTACTTTTCAAGTAACACAA 1140	Qy
Qy	1144 CTCTTCCATCAGGCCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTGTGAC 1203	Db
Db	1141 CTCTTCCATCAGGCCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTGTGAC 1200	Qy
Qy	1204 CCCAAACCATCTCTCAAAGCATTAATATCAATCATGCGCTGTATGTTTAAATCAAGCAG 1263	Db
Db	1201 CCCAAACCATCTCTCAAAGCATTAATATCAATCATGCGCTGTATGTTTAAATCAAGCAG 1260	Qy
Qy	1264 AAGCATGTTTTTATGTTGTACAAAAGAGTTGTTATGGTGGGATGAGGATATAGAC 1323	Db
Db	1261 AAGCATGTTTTTATGTTGTACAAAAGAGTTGTTATGGTGGGATGAGGATATAGAC 1320	Qy
Qy	1324 CATGCATGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAATGGGACAGGAGGACTGT 1383	Db
Db	1321 CATGCATGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAATGGGACAGGAGGACTGT 1380	Qy
Qy	1384 GAACAAGACACCCCTAATAATGGGTTGATGTCTCAAGTAGCAATCTTCTGGAACGCA 1443	Db
Db	1381 GAACAAGACACCCCTAATAATGGGTTGATGTCTCAAGTAGCAATCTTCTGGAACGCA 1440	Qy
Qy	1444 CTCCTTTAAGGAAGTCCCTAAATTTAGAAACACCCACAACTTCACATATCATTAATAGCA 1503	Db
Db	1441 CTCCTTTAAGGAAGTCCCTAAATTTAGAAACACCCACAACTTCACATATCATTAATAGCA 1500	Qy
Qy	1504 ACAATTTGGAAGCAAGTCTCTCAATGCTTGGGAGAGGAAATCTATGCTCTCGTGGG 1563	Db
Db	1501 ACAATTTGGAAGCAAGTCTCTCAATGCTTGGGAGAGGAAATCTATTTGCTCTCGTGGG 1560	

QY	1564	TCCTTCATCTCAGAAATGCCAATCAGGTCAGGTTTGCTACAGTTTGTGATGTTGAT	1623
Db	1561	TCCTTCATCTCAGAAATGCCAATCAGGTCAGGTTTGCTACATTTTGTGATGTTGAT	1620
QY	1624	GCTTCTCCCAAGGTATATTAATATATATAGAGAGTTGTGACAAAACAGAAATGATAAGC	1683
Db	1621	GCTTCTCCCAAGGTATATTAATATATATAGAGAGTTGTGACAAAACAGAAATGATAAGC	1680
QY	1684	TGGGAACCGTGGCACACGCTCATAGTCTAGCTGCTGGAGGCTTGAGGAGGAGATGG	1743
Db	1681	TGGGAACCGTGGCACACGCTCATAGTCTAGCTGCTGGAGGCTTGAGGAGGAGATGG	1740
QY	1744	CTTGAACACAGGTGTTCAAGGCCAGGCTGGGCAACATACAAGATCCTGCTCTCAAAA	1803
Db	1741	CTTGAACACAGGTGTTCAAGGCCAGGCTGGGCAACATACAAGATCCTGCTCTCAAAA	1800
QY	1804	AAAAAANAANAAGAGAGAGAGCGCGGGCTGGTGCTCAGCCCTGTAATCCAG	1863
Db	1801	AAAAAANAANAAGAGAGAGAGCGCGGGCTGGTGCTCAGCCCTGTAATCCAG	1860
QY	1864	CACTTTGGAGCGGAGCGGCGGATCACTCTGCTCAGAGTTCAGACACGAGCTGGC	1923
Db	1861	CACTTTGGAGCGGAGCGGCGGATCACTCTGCTCAGAGTTCAGACACGAGCTGGC	1920
QY	1924	CAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATTAGCCAGGCGTGTAGCAGG	1983
Db	1921	CAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATTAGCCAGGCGTGTAGCAGG	1980
QY	1984	CACCTGTAATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGAGGT	2043
Db	1981	CACCTGTAATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGAGGT	2040
QY	2044	GGAGGTGAGTAAAGCTGAGATCGTCCGTTGCACCTCCAGCCTGGCGCAAGCAAGA	2103
Db	2041	GGAGGTGAGTAAAGCTGAGATCGTCCGTTGCACCTCCAGCCTGGCGCAAGCAAGA	2100
QY	2104	CTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2163
Db	2101	CTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2160
QY	2164	GAGAAGATGGGAAGCATTCAGAGAAATTTGCTTTATCCAAACAAATGTAAAGAGCC	2223
Db	2161	GAGAAGATGGGAAGCATTCAGAGAAATTTGCTTTATCCAAACAAATGTAAAGAGCC	2220
QY	2224	AATAAGGATCCCTATTGTCCTTTTGGTGCTTATTTGTCCTTACCACTGCTTTGAC	2283
Db	2221	AATAAGGATCCCTATTGTCCTTTTGGTGCTTATTTGTCCTTACCACTGCTTTGAC	2280
QY	2284	AGTGAGAAATATTTCAGATAACCATATCCCTGTGCGCTTATTACCTAGCAACCTTGC	2343
Db	2281	AGTGAGAAATATTTCAGATAACCATATCCCTGTGCGCTTATTACCTAGCAACCTTGC	2340
QY	2344	AATGAGATGAGCAGATCCACAGAAACATTGAATGCACAACTGCTTATTTAACTTA	2403
Db	2341	AATGAGATGAGCAGATCCACAGAAACATTGAATGCACAACTGCTTATTTAACTTA	2400
QY	2404	TTGTACATAGTTTGTAAAGAGCTTAAATTTGTTACTTCATCTATTCATTTATTTTA	2463
Db	2401	TTGTACATAGTTTGTAAAGAGCTTAAATTTGTTACTTCATCTATTCATTTATTTTA	2460
QY	2464	TATTATTTTTCGCTCAATGATTTTATTAAACATGATTTTCTTTCTGATATATTGAAAT	2523
Db	2461	TATTATTTTTCGCTCAATGATTTTATTAAACATGATTTTCTTTCTGATATATTGAAAT	2520
QY	2524	GGAGTCTCAAGCTTCATAATTTATAACTTTTGAAGATGATCTAATAACAACGATGTA	2583
Db	2521	GGAGTCTCAAGCTTCATAATTTTATAACTTTTGAAGATGATCTAATAACAACGATGTA	2580
QY	2584	ATTGTRACATTGAGTAAAGGCTCTACGAGCCATTTCCTTCTGATTTTGTAACTTTT	2643
Db	2581	ATTGTRACATTGAGTAAAGGCTCTACGAGCCATTTCCTTCTGATTTTGTAACTTTT	2640

QY	2644	ATGACAGCAAAATTTGCTTCTGGCTCACTTCAATCAAGTTAAATAAATGATAAATAATTTT	2703
Db	2641	ATGACAGCAAAATTTGCTTCTGGCTCACTTCAATCAAGTTAAATAAATGATAAATAATTTT	2700
QY	2704	GGAAGCTGTGAAGATAAATAATACCAATAAATAATATATAAAGTGATTTATGAAGTTAA	2763
Db	2701	GGAAGCTGTGAAGATAAATAATACCAATAAATAATATATAAAGTGATTTATGAAGTTAA	2760
QY	2764	AATAAAAAATCAGTATGATGGAATAAACTTG	2794
Db	2761	AATAAAAAATCAGTATGATGGAATAAACTTG	2791
RESULT 2			
AB040430			
LOCUS			
DEFINITION	Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.	11204 bp	DNA linear PRI 03-OCT-2000
ACCESSION	AB040430		
VERSION	AB040430.1	GI:9988407	
KEYWORDS	AID: activation-induced cytidine deaminase.		
SOURCE	Homo sapiens DNA.		
ORGANISM	Homo sapiens		
REFERENCE	1 (sites)		
AUTHORS	Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.		
TITLE	Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene		
JOURNAL	Genomics 68 (1), 85-88 (2000)		
MEDLINE	20408890		
REFERENCE	2 (sites)		
AUTHORS	Rev, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Genery, A., Tezcan, I., Ersoy, K., Kayser, H., Ugazio, A.G., Brousse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A.		
TITLE	Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)		
JOURNAL	Cell 102 (5), 565-575 (2000)		
MEDLINE	20460541		
REFERENCE	3 (bases 1 to 11204)		
AUTHORS	Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.		
TITLE	Direct Submision		
JOURNAL	Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine: Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjoem@our.med.kyoto-u.ac.jp, Tel: 81-75-753-4371(ex.4371), Fax: 81-75-753-4388)		
FEATURES	Location/Qualifiers		
source	1. 11204		
gene	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)		
	/gene="AID"		
	join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)		
	/gene="AID"		
	/codon_start=1		
	/product="activation-induced cytidine deaminase"		
	/protein_id="BAB12720.1"		
	/db_xref="GI:9988408"		
	/translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL		
	NFLGRNKGCHVELLFRLYISDWLDPGRCYRVFTWSPCYDCARHVAFLRGNP		
	DFSLRFTARLYFCEKRAPEGLRLRHAGVOIATMTKDFYCNWTFVENHERTEK		
	AWGLHENSRLSRQLRLLPLLYEVDLDLDAFTLGL"		
BASE COUNT	3305 a 2273 c 2373 g 3253 t		
ORIGIN			
Query Match	77.2%	Score 2174.6;	DB 9; Length 11204;
Best Local Similarity	99.8%	Pred. No. 0;	
Matches 2177; Conservative	0;	Mismatches 4;	Indels 0; Gaps 0;
QY	617	CTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCGTACTTTGGGACTTTGA	676
Db	617	CTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCGTACTTTGGGACTTTGA	676

Db	8950	CTACAGCCCTGTATGAGTTTGATGACCTTACGAGCGCATTCGTA	1757	GTTCAGGCCAGCCTGGGCAACATTAACAGATCCTCTCTCAAAAAA	1816
Qy	677	TAGCAACTTCAGGAATGTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATAAAA	10090	GTTCAGGCCAGCCTGGGCAACATTAACAGATCCTCTCTCAAAAAA	10149
Db	9010	TAGCAACTTCAGGAATGTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATAAAA	1817	AAGAAAGAGAGAGGCCCGGGCGTGGTCTCACGGCTCTAATCCAGACACTTTGGAGGC	1876
Qy	737	AACAGTCTCTCAAGTCTCTCTGTTTTATCTTCAACTCTCACCTTTCTTAGAGTTTACA	10150	AAGAAAGAGAGAGGCCCGGGCGTGGTCTCACGGCTCTAATCCAGACACTTTGGAGGC	10209
Db	9070	AACAGTCTCTCAAGTCTCTCTGTTTTATCTTCAACTCTCACCTTTCTTAGAGTTTACA	1877	CGAGCCGGCGGATCACCTGTGGTCAGAGTTTGGAGACAGCCTGGCCCAACATGCAAAA	1936
Qy	797	GAATAAATPATTATATAGACTCTTTAAAAAGATCTATGTCTGAAAAATAGAGAAGAAC	10210	CGAGCCGGCGGATCACCTGTGGTCAGAGTTTGGAGACAGCCTGGCCCAACATGCAAAA	10269
Db	9130	GAATAAATPATTATATAGACTCTTTAAAAAGATCTATGTCTGAAAAATAGAGAAGAAC	1937	CCCCGTCTGTACTCAAAATTTAGCCAGGCGTGTAGCAGGCACCTGTAAATCCC	1996
Qy	857	ACAGTCTGGCCAGGAGCGTCTGCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA	10270	CCCCGTCTGTACTCAAAATTTAGCCAGGCGTGTAGCAGGCACCTGTAAATCCC	10329
Db	9190	ACAGTCTGGCCAGGAGCGTCTGCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA	1997	AGCTACTTTGGGAGCGTGAGCGAGGAGAAATCGCTTGAACCCAGGAGTGGACCTA	2056
Qy	917	CTGGGAATACAGAACTGCAGACCTGGGAGCATCCTAAAGATGTCAGAGTTTCTATGA	10330	AGCTACTTTGGGAGCGTGAGCGAGGAGAAATCGCTTGAACCCAGGAGTGGACCTA	2056
Db	9250	CTGGGAATACAGAACTGCAGACCTGGGAGCATCCTAAAGATGTCAGAGTTTCTATGA	2057	AGCTGAGATCGTCCGTTGCACTCCAGCTGGGCGACAAGACAGACTCTGTCTCAGAA	2116
Qy	977	CTTTTAGGTAGGATGAGACGAAAGTAGATCCTTAAAAAGCATGGTGAGAGATCAAAATG	10390	AGCTGAGATCGTCCGTTGCACTCCAGCTGGGCGACAAGACAGACTCTGTCTCAGAA	2116
Db	9310	CTTTTAGGTAGGATGAGACGAAAGTAGATCCTTAAAAAGCATGGTGAGAGATCAAAATG	2117	AAAAAAG	2176
Qy	1037	TTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTTAAACAGTGGTGTAGTATA	10450	AAAAAAG	2176
Db	9370	TTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTTAAACAGTGGTGTAGTATA	2177	AAGCATTGCAAGGAAATTTGCTTTATCCAAACAAATTTAAGAGGCAATTAAGGATGCC	2236
Qy	1097	GATTTTCTATTTCTTTCCCTTACGTTTCAAGTACACAAACTCTTCCATCAGG	10510	AAGCATTGCAAGGAAATTTGCTTTATCCAAACAAATTTAAGAGGCAATTAAGGATGCC	2236
Db	9430	GATTTTCTATTTCTTTCCCTTACGTTTCAAGTACACAAACTCTTCCATCAGG	2237	TATTTGCTCTCTTTGGTGTCTATTTGTCCTTAACTGCTTTGACAGTGGAGGAAATA	2296
Qy	1157	CCATGATCTATAGACCTCCTTAATGAGAGTATCTGGGTGATTTGACCCCAACCACTCTC	10570	TATTTGCTCTCTTTGGTGTCTATTTGTCCTTAACTGCTTTGACAGTGGAGGAAATA	2296
Db	9490	CCATGATCTATAGACCTCCTTAATGAGAGTATCTGGGTGATTTGACCCCAACCACTCTC	2297	TTCAGAAATACCAATPATCCCTGTGCCGTTATTACCTAGCAACCCCTTGAACATGAAGT	2356
Qy	1217	TCCAAAGCATTATATCCATCATCGCTGTATGTTTAAATCAGCAGAGCATGTTTTTA	10630	TTCAGAAATACCAATPATCCCTGTGCCGTTATTACCTAGCAACCCCTTGAACATGAAGT	2356
Db	9550	TCCAAAGCATTATATCCATCATCGCTGTATGTTTAAATCAGCAGAGCATGTTTTTA	2357	AGATCCACAGGAAACCTTGAATGCACAACCTGTCTTATTTAACTTTATGCTACATAAGTT	2416
Qy	1277	TGTTTGTACAAAAGAGATGTTTATGGGTGGGATGGAGGTATAGACCATGCGTGCAC	10690	AGATCCACAGGAAACCTTGAATGCACAACCTGTCTTATTTAACTTTATGCTACATAAGTT	2416
Db	9610	TGTTTGTACAAAAGAGATGTTTATGGGTGGGATGGAGGTATAGACCATGCGTGCAC	2417	TGTAAGAGAGTTAAAAATTTGTTACTTCATGTATTTCATTTATTTATTTATTTTCGCT	2476
Qy	1337	CTTCAAGCTACTTTAATAAGGATCTTAAATGGCAGGAGGACTGTGAACAAGACACCC	10750	TGTAAGAGAGTTAAAAATTTGTTACTTCATGTATTTCATTTATTTATTTATTTTCGCT	2476
Db	9670	CTTCAAGCTACTTTAATAAGGATCTTAAATGGCAGGAGGACTGTGAACAAGACACCC	2477	CTAATGATTTTTTATTAAACATGATTCCTTTCTGATATATTGAAATGGAGTCTCAAGC	2536
Qy	1397	TAATAATGGGTGATGCTCAAGTACCAATCTCTGGAACGCAAACTCTTTTAAGGAA	10810	CTAATGATTTTTTATTAAACATGATTCCTTTCTGATATATTGAAATGGAGTCTCAAGC	2536
Db	9730	TAATAATGGGTGATGCTCAAGTACCAATCTCTGGAACGCAAACTCTTTTAAGGAA	2537	TTCAATAATTTATTAACCTTTTGAAGATGATTAACAACGCTATGTAATTTGTAACATTC	2596
Qy	1457	GTCCCTAATTTAGAAACACCCACAACTTCACATATCATAAATAGCAAACTTTGGAAG	10870	TTCAATAATTTATTAACCTTTTGAAGATGATTAACAACGCTATGTAATTTGTAACATTC	2596
Db	9790	GTCCCTAATTTAGAAACACCCACAACTTCACATATCATAAATAGCAAACTTTGGAAG	2597	AGTAATGGTCTACGAGCCATTTCTCTGATTTTTAGTAACTTTTATGACAGCAAAAT	2656
Qy	1517	AAGTTCCTGAATTTGGGAGAGGAAATCTAATGGCTCTCGTGGGTCTCTTCATCTCA	10930	AGTAATGGTCTACGAGCCATTTCTCTGATTTTTAGTAACTTTTATGACAGCAAAAT	2656
Db	9850	AAGTTCCTGAATTTGGGAGAGGAAATCTAATGGCTCTCGTGGGTCTCTTCATCTCA	2657	TGCTTCTGGCTCACTTTCAATAGTTAAATAAATGATAAATAAATTTTGGAGCTGTGAAG	2716
Qy	1577	GAATGCAATCAGGTCAAGTTTGTACATTTTGTATGTTGATGCTTCTCCCAAG	10990	TGCTTCTGGCTCACTTTCAATAGTTAAATAAATGATAAATAAATTTTGGAGCTGTGAAG	2716
Db	9910	GAATGCAATCAGGTCAAGTTTGTACATTTTGTATGTTGATGCTTCTCCCAAG	2717	ATAAAAACCAATAAATAAATAAAGTGATTTATCAAGCTTAAATAAATAAATAAATCAG	2776
Qy	1637	GTATATTAACATATAAGAGTTGTGACAAAAAGAAATGATAAAGCTGCGAACCTGGC	11050	ATAAAAACCAATAAATAAATAAAGTGATTTATCAAGCTTAAATAAATAAATAAATCAG	2776
Db	9970	GTATATTAACATATAAGAGTTGTGACAAAAAGAAATGATAAAGCTGCGAACCTGGC	2777	TATGATGGAATAAACTTTGAAA	2797
Qy	1697	ACAGGCTCATAGTTCTAGTCTTGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	11110	TATGATGGAATAAACTTTGAAA	2797
Db	10030	ACAGGCTCATAGTTCTAGTCTTGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT			

repeat_region	3296..3328	/rpt_family=" (TTTC)n"	Qy	797	GAATAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAATAGAGAGAAC	856
repeat_region	4097..4249	/rpt_family="AluSq"	Db	43673	GAATAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAATAGAGAGAAC	43732
repeat_region	4360..4654	/rpt_family="AluY"	Qy	857	ACAGGTCTGGCCAGGAGCTGCTGCAATTTGGTGAGATTTTGAATGCAACATTTGCCCTTA	916
repeat_region	complement(5140..5262)	/rpt_family="FLAM_C"	Db	43733	ACAGGTCTGGCCAGGAGCTGCTGCAATTTGGTGAGATTTTGAATGCAACATTTGCCCTTA	43792
repeat_region	5275..5304	/rpt_family="AT-rich"	Qy	917	CTGGGAATAACAGAACTGCGAGGACCTGGAGACCTCTAAAGTGTCAACGTTTTTCTATGA	976
repeat_region	5305..5563	/rpt_family="AluJo"	Db	43793	CTGGGAATAACAGAACTGCGAGGACCTGGAGACCTCTAAAGTGTCAACGTTTTTCTATGA	43852
repeat_region	5564..5599	/rpt_family="AT-rich"	Qy	977	CTTTTAGGTAGGTAGGATGAGAGAGAGTGTCTTAAAGAGCATGGTGAGAGATCAAAATG	1036
repeat_region	5644..5733	/rpt_family="MSTD"	Db	43853	CTTTTAGGTAGGTAGGATGAGAGAGAGTGTCTTAAAGAGCATGGTGAGAGATCAAAATG	43912
repeat_region	5734..6012	/rpt_family="AluSq"	Qy	1037	TTTTTATATCAACATCTCTTTATTTATTTGATTTTCAAGTGTCAACGTTGGTGTAGTATGA	1096
repeat_region	6020..6336	/rpt_family="AluSx"	Db	43913	TTTTTATATCAACATCTCTTTATTTATTTGATTTTCAAGTGTCAACGTTGGTGTAGTATGA	43972
repeat_region	6337..6629	/rpt_family="MSTD"	Qy	1097	GATTTTCTATTTCTTTCCCTTTGACGTTTACTTTCAAGTGTCAACGTTCTTCCATCAGG	1156
repeat_region	6630..6650	/rpt_family="(CAA)n"	Db	43973	GATTTTCTATTTCTTTCCCTTTGACGTTTACTTTCAAGTGTCAACGTTCTTCCATCAGG	44032
repeat_region	complement(6711..6817)	/rpt_family="MIR"	Qy	1157	CCATGATCTATAGGACCTCTCTTAATGAGAGTATCTGGGTGATGTGACCCCAACCATCTCT	1216
repeat_region	complement(7033..7327)	/rpt_family="AluSx"	Db	44033	CCATGATCTATAGGACCTCTCTTAATGAGAGTATCTGGGTGATGTGACCCCAACCATCTCT	44092
repeat_region	7607..7745	/rpt_family="FLAM_A"	Qy	1217	TCCAAGCATTAATATCAACATCATGCGCTGTATCTTTTAATCAGCAGAGCATCTTTTATA	1276
repeat_region	7980..8047	/rpt_family="MIR"	Db	44093	TCCAAGCATTAATATCAACATCATGCGCTGTATCTTTTAATCAGCAGAGCATCTTTTATA	44152
repeat_region	complement(8050..8549)	/rpt_family="LTR47A"	Qy	1277	TGTTTGTACAAAAGAAGATTGTTATGGTGGGGATGGAGGTATAGACCATCATCTGGTCAC	1336
repeat_region	8567..8698	/rpt_family="MIR"	Db	44153	TGTTTGTACAAAAGAAGATTGTTATGGTGGGGATGGAGGTATAGACCATCATCTGGTCAC	44212
repeat_region	complement(8815..9101)	/rpt_family="AluSg"	Qy	1337	CTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGGACTGTGAACAACACACACC	1396
repeat_region	complement(11227..12175)	/rpt_family="LTR5"	Db	44213	CTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGGACTGTGAACAACACACACC	44272
STS	12625..12900	/standard_name="87432"	Qy	1397	TAATAAGTGGTGTGATCTGAAAGTGTGAAAGTCTTCTGAAACGCAAACTCTTTTAAGGAA	1456
repeat_region	complement(13547..13664)	/rpt_family="LMB5"	Db	44273	TAATAAGTGGTGTGATCTGAAAGTGTGAAAGTCTTCTGAAACGCAAACTCTTTTAAGGAA	44332
repeat_region	complement(13665..13979)	/rpt_family="AluY"	Qy	1457	GTCCCTAATTTAGAAAACACCCACAACTTCACATATCATATTAATAGCAAACTTTGGAAG	1516
repeat_region	complement(13980..14193)	/rpt_family="LMB5"	Db	44333	GTCCCTAATTTAGAAAACACCCACAACTTCACATATCATATTAATAGCAAACTTTGGAAG	44392
repeat_region	14622..14924	/rpt_family="AluY"	Qy	1517	AAGTTGCTTGAATGTGGGGAGAGAAATCTATTTGGCTCTCGTGGGTCTCTTCATCTCA	1576
repeat_region	complement(15813..17136)	/rpt_family="LIP48"	Db	44393	AAGTTGCTTGAATGTGGGGAGAGAAATCTATTTGGCTCTCGTGGGTCTCTTCATCTCA	44452
repeat_region	complement(17272..17395)	/rpt_family="FLAM_C"	Qy	1577	GAATGCCAATCAGGTCAAGGTTTGTACATTTTGTATGTGTGTGTCTCTCCCAAG	1636
repeat_region	17426..17575	/rpt_family="MIR"	Db	44453	GAATGCCAATCAGGTCAAGGTTTGTACATTTTGTATGTGTGTGTCTCTCCCAAG	44512
repeat_region	19021..19314	/rpt_family="MIR"	Qy	1637	GTATATTAACATATATAAGAGAGTTGTGACAAAACAGAAATGATAAGCTGCGAACCGTGGC	1696
Query Match	77.2%; Score 2174.6; DB 9; Length 71132;		Db	44513	GTATATTAACATATATAAGAGAGTTGTGACAAAACAGAAATGATAAGCTGCGAACCGTGGC	44572
Best Local Similarity	99.8%; Pred. No. 0;		Qy	1697	ACAGCTCATAGTTCTAGCTCTTTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	1756
Matches 2177; Conservative	0; Mismatches 4; Indels 0; Gaps 0;		Db	44573	ACAGCTCATAGTTCTAGCTCTTTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	44632
Qy	617	CTTTTGGCCCTGTATGAGTTGATGACTTACGAGCAGCATTTCCGTACTTTGGGACTTTGA	676			
Db	43493	CTACAGCCCTGTATGAGTTGATGACTTACGAGCAGCATTTCCGTACTTTGGGACTTTGA	43552			
Qy	677	TAGCAACTTCAGGATCTCACACGATGAATATCTCTGCTGAGACAGTGGATAAA	736			
Db	43553	TAGCAACTTCAGGATCTCACACGATGAATATCTCTGCTGAGACAGTGGATAAA	43612			
Qy	737	AACAGTCTCTCAAGTCTCTCTGTTTTTATTTCTTCAACTCTCACCTTTCTAGAGTTTACA	796			
Db	43613	AACAGTCTCTCAAGTCTCTCTGTTTTTATTTCTTCAACTCTCACCTTTCTAGAGTTTACA	43672			

[illegible]

RESULT 4	
BC006296	
LOCUS	1837 bp linear mRNA
DEFINITION	Homo sapiens, activation-induced cytidine deaminase, clone MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION	BC006296
VERSION	BC006296.1 GI:13623400
KEYWORDS	MGC.

Qy	181	GAGCGGTGACAGTGCTACATCCTTTTTCACCTGGACTTTGGTTATCTTCCCAATAAGAACGG	240
Db	187	GAGCGGTGACAGTGCTACATCCTTTTTCACCTGGACTTTGGTTATCTTCCCAATAAGAACGG	246
Qy	241	CTGCCAGTGGAAATGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTTGCCG	300
Db	247	CTGCCAGTGGAAATGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTTGCCG	306
Qy	301	CTGCTACCGGTCACCTGGTTTCACTCTCGAGCCCTGCTAGGACTGTGCCGACATGT	360
Db	307	CTGCTACCGGTCACCTGGTTTCACTCTCGAGCCCTGCTAGGACTGTGCCGACATGT	366
Qy	361	GGCGGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGGCTCTTA	420
Db	367	GGCGGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGGCTCTTA	426
Qy	421	CTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGTGCACCGCGCGGGGT	480
Db	427	CTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGTGCACCGCGCGGGGT	486
Qy	481	GCAAAATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAA	540
Db	487	GCAAAATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAA	546
Qy	541	CCATGAAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAG	600
Db	547	CCATGAAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAG	606
Qy	601	ACAGCTTCGGCGCATCCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCCG	660
Db	607	ACAGCTTCGGCGCATCCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCCG	666
Qy	661	TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTG	720
Db	667	TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTG	726
Qy	721	AAGACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCAC	780
Db	727	AAGACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCAC	786
Qy	781	TTTCTTAGAGTTTACAGAAAAATATTTATATAGACTCTTTAAAAAGATCTATGCTCTG	840
Db	787	TTTCTTAGAGTTTACAGAAAAATATTTATATAGACTCTTTAAAAAGATCTATGCTCTG	846
Qy	841	AAAAATAGAGAAGAACACAGCTCTGGCCAGGAGCTGCTCAATTTGGTGAGTTTGAAT	900
Db	847	AAAAATAGAGAAGAACACAGCTCTGGCCAGGAGCTGCTCAATTTGGTGAGTTTGAAT	906
Qy	901	GCAACATTTGCCCTACTCTGGGAATACAGAACTGCAGGACCTGGGAGCATCTCTAAAGTGT	960
Db	907	GCAACATTTGCCCTACTCTGGGAATACAGAACTGCAGGACCTGGGAGCATCTCTAAAGTGT	966
Qy	961	CAACGTTTTTCTATGACTTTTAGTGTAGGATGAGAGCAGAGGTAGATCCTTAAAAAGCATG	1020
Db	967	CAACGTTTTTCTATGACTTTTAGTGTAGGATGAGAGCAGAGGTAGATCCTTAAAAAGCATG	1026
Qy	1021	GTGAGAGATCAAAATGTTTTATATCAACATCCTTTTATTTATTTGATTCATTTGAGTTAAC	1080
Db	1027	GTGAGAGATCAAAATGTTTTATATCAACATCCTTTTATTTATTTGATTCATTTGAGTTAAC	1086
Qy	1081	AGTGGTGTAGTATAGATTTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACAC	1140
Db	1087	AGTGGTGTAGTATAGATTTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACAC	1146
Qy	1141	AAACTCTTCCATCAGGCGCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTGT	1200
Db	1147	AAACTCTTCCATCAGGCGCATGATCTATAGGACCTCTCTTAATGAGAGTATCTGGGTGATTGT	1206
Qy	1201	GACCCCAAAACATCTCTCCAAAGCATTAATATCAATCATCGCTGTATGTTTTTAATCAG	1260
Db	1207	GACCCCAAAACATCTCTCCAAAGCATTAATATCAATCATCGCTGTATGTTTTTAATCAG	1266
Qy	1261	CAGAAGCATGTTTTTTATGTTTGTACAAAAGAAATGTTTATGGTGGGGATGGAGGTATA	1320
Db	1267	CAGAAGCATGTTTTTTATGTTTGTACAAAAGAAATGTTTATGGTGGGGATGGAGGTATA	1326
Qy	1321	GACCATGCATGGTCCACCTTCAAGCTACTTTAATAAAGATCTTAAAAATGGCAGGAGGAC	1380
Db	1327	GACCATGCATGGTCCACCTTCAAGCTACTTTAATAAAGATCTTAAAAATGGCAGGAGGAC	1386
Qy	1381	TGTGAACAAGACACCCCTAAATAATGGGTTGATGCTGAAGTAGCAAAATCTTCTGGAAACGC	1440
Db	1387	TGTGAACAAGACACCCCTAAATAATGGGTTGATGCTGAAGTAGCAAAATCTTCTGGAAACGC	1446
Qy	1441	AAACTCTTTTAAAGGAAGTCCCTTAATTTAGAAACACCCACAAACTTCACATATCATATAA	1500
Db	1447	AAACTCTTTTAAAGGAAGTCCCTTAATTTAGAAACACCCACAAACTTCACATATCATATAA	1506
Qy	1501	GCAAAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAAAATCTATTGGCTCTCGT	1560
Db	1507	GCAAAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAAAATCTATTGGCTCTCGT	1566
Qy	1561	GGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGTCTACATTTTGTATGTGTGT	1620
Db	1567	GGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGTCTACATTTTGTATGTGTGT	1626
Qy	1621	GATGCTTCTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAAAACAGAAATGATAA	1680
Db	1627	GATGCTTCTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAAAACAGAAATGATAA	1686
Qy	1681	AGCTGGGAACCGTGGCACGCTCATAGTTCTAGCTGCTTGGGAGGTTTGAGGAGGGAGGA	1740
Db	1687	AGCTGGGAACCGTGGCACGCTCATAGTTCTAGCTGCTTGGGAGGTTTGAGGAGGGAGGA	1746
Qy	1741	TGCTTTGAACACACAGGTTTCAAGCCAGCCTGGGCAACATAAACAGATCTCTCTCTCAA	1800
Db	1747	TGCTTTGAACACACAGGTTTCAAGCCAGCCTGGGCAACATAAACAGATCTCTCTCTCAA	1806
Qy	1801	AAAAAAAAAAAAAAAAAAAAAGAGAGAGA	1828
Db	1807	AAAAAAAAAAAAAAAAAAAAAGAGAGAGA	1834
RESULT 5			
AF132979			
LOCUS	2440 bp mRNA linear ROD 05-AUG-1999		
DEFINITION	Mus musculus activation-induced cytidine deaminase (AID) mRNA,		
complete cds.			
ACCESSION	AF132979		
VERSION	AF132979.1 GI:5305727		
KEYWORDS	Mus musculus.		
SOURCE	Mus musculus		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 2440)		
TITLE	Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K.,		
	Davidson, N.O. and Honjo, T.		
	Specific expression of activation-induced cytidine deaminase (AID),		
	a novel member of the RNA-editing deaminase family in germinal		
	center B cells		
JOURNAL	J. Biol. Chem. 274 (26), 18470-18476 (1999)		
MEDLINE	99303612		
PUBMED	10373455		
REFERENCE	2 (bases 1 to 2440)		
AUTHORS	Muramatsu, M. and Honjo, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-NAR-1999) Medical Chemistry, Kyoto University, Konoe		
	Yoshida Sakyo-ku, Kyoto 606-8501, Japan		
FEATURES	Location/Qualifiers		
source	1..2440		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
gene	1..2440		
	/gene="AID"		
CDS	93..689		

		/gene="AID"		706 a 546 c 551 g 636 t 1 others	
		/note="APOBEC-1 homologue"			
		/codon_start=1			
		/product="activation-induced cytidine deaminase"			
		/protein_id="AAD41793.1"			
		/db_xref="gi:5305728"			
		/translation="MDSLIMKQKKFLYHFKNVRWAKGRHETLYLCVYVKKRRDSATSCSL DFGLHRSKSGHVELFLRYISDMDLDPRCYRVYVFTWSPCYDCARHVAEFLRWNP NLSRIETARLYFCEDRKAEPGLRHRAGVQIGIMTFKDYFCVWNTFVENRETEFK AWEGLHNSVRLTQLRILLPLLYEVDLDRDAFRLGF"			
BASE COUNT		706 a 546 c 551 g 636 t 1 others			
ORIGIN					
Query Match		21.4%; Score 603.4; DB 10; Length 2440;			
Best Local Similarity		69.4%; Pred. No. 1.2e-108;			
Matches 1006; Conservative		0; Mismatches 351; Indels 92; Gaps 10;			
QY	4	GAACATCATTAATTAAGTGGAGATTTTTCGGCCGTGAGACTTGCGAGGGAGCAAG-AAG	62		
DB	16	GAAGCAGCCTTGCTTGAAGCAAGCTTCTCTTGGCCTAAGACTTTGAGGGAGTCAGAAAG	75		
QY	63	ACACTCTGCACCACTATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAAT	122		
DB	76	TCACCTGGAGCCGATATGACAGCCTCTTGATGAACCAAGAAAGTTCTTTACCAAT	135		
QY	123	TCAAAAATGTCGGCTGGCTAAGGTCGGCGGTGAGACCTACTGTGCTACGTAGTGAAGA	182		
DB	136	TCAAAAATGTCGGCTGGCGCAAGGAGCGCATGAGACCTACTCTGCTACGTGGTGAAGA	195		
QY	183	GGCTGACAGTGTACATCCTTTTACCTGGAGCTTTGGTTATCTTCGCAATAGAAGCGCT	242		
DB	196	GGAGAGATGAGTGCACCTCTGCTCACTGGAGCTTTCGGCCACCTTTCGCAACAGTCTGGCT	255		
QY	243	GCCAGTGGAAATGCTCTCTCCGTACACTTCGACTTCGAGCTGGACCTAGACCTGCGCGCT	302		
DB	256	GCCAGTGGAAATGCTCTCTCCGTACACTTCGACTTCGAGCTGGACCTAGACCTGCGCGCT	315		
QY	303	GCTACCGCGTCACTGGTTCACCTCCTCGAGCCCTGCTACGACTGTGCCCGACATGTGG	362		
DB	316	GTTACCGCGTCACTGGTTCACCTCCTCGAGCCCTGCTACGACTGTGCCCGACATGTGG	375		
QY	363	CGGACTTCTCGGAGGGAACCCCACTCACTGCTGAGGATCTTACCGGGCGCCTCTACT	422		
DB	376	CTGAGTTCTTGAGATGAGAACCCCTAACCTCAGCTGAGGATTTTTCACCGCGCCTCTACT	435		
QY	423	TCGTGTGAGACCGCAAGCTGAGCCGAGGGCTCGCGGCTGCACCGCGCGGGGTGC	482		
DB	436	TCTGTGAAGCCGCAAGCTGAGCCTGAGGGCTCGGAGACTGCACCGCGCTGGGTGC	495		
QY	483	AAATAGCCATCATGACCTTCAAGATATTTTACTGCTGGAATATCTTTGTAGAAAACC	542		
DB	496	AGATCGGATCATGACCTTCAAGACTATTTTACTGCTGGAATATCTTTGTAGAAAATC	555		
QY	543	ATGAAGAATCTTCAAGCCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGAC	602		
DB	556	GTGAAGAATCTTCAAGCCTGGGAAGGCTGCATGAATAATCTCTCGGCTTAACAGAC	615		
QY	603	AGCTTCGGCGATCTTTTGGCCCTGATGAGTTGATGACTTACGAGCAGCATTCGTA	662		
DB	616	AACTTCGGCGATCTTTTGGCCCTGATGAGTTGATGACTTACGAGCAGCATTCGTA	675		
QY	663	CTTTGGGACTTTGATAGCACTTCCAGGAATGTCACACGATGAATATCTCTCTGAA	722		
DB	676	TGTTGGGATTTTGAAGAACAACCTCTGGAATGTCACAGTGATGAATTTCTI--CTGAA	732		
QY	723	GACAGTGATGAATAAAGAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTT	782		
DB	733	GAGATGGATGAATAAAGCAACCTTCA--ACTACATGTTTCTTCTTAAAGTACTCACTT	790		
QY	783	TCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGACTATGCTTGA	842		
DB	791	TTATAAGTGTAGGGGAAA-----TTATATGACTTTTAAAAAATACTTGAAGTCAC	843		
QY	843	AATAGAGGAAGAACACAGGTCTGGCCAGGGACGTGCTGCAATTTGGTGCAGTTTGAATGC	902		
DB	844	AGGACCGCAGAGCAATGATGTAAC-----TGAGCTTGTCTGTGC	882		
QY	903	AACATTGTCCTTACT-GGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAGATGTC	961		
DB	883	AACATCGGCATCTACTGGGGAACAGCAATACTCCAGACTTTGGGTCTGGAATGATGTC	942		
QY	962	AACGTGTTTCTATGACTTTTAGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGACGG	1021		
DB	943	TTTTTTTTCACAGCAT-----GGAAGACGATAT	971		
QY	1022	TGAGAGGATCAATGTTTTTATATCAACATCCTTTATTATTGATTCATTTGAGTAAACA	1081		
DB	972	GGAGAGGACCAACACAGTTTGTACACCCACCTGTGTTCTTGAATTCATTTGAATCTCA	1031		
QY	1082	GTGGTGTAGTATAGATTTTCTTATTTTCCCTTACCTTACCTTACCTTACCTTACCTT	1141		
DB	1032	GGGTATCAGTGACGATTTCTTATTTTCCCTTACCTTACCTTACCTTACCTTACCTT	1084		
QY	1142	AACCTCTCCATCAGGCCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATGTG	1201		
DB	1085	---GGTCTTCTCACAAGTCAACGGGCTCTCTACAGTCTCTCTCTGAGCAAT----	1137		
QY	1202	ACCCCAACCATCTCTCCAAAGCATTAATATCAATCAATCAATCAATCAATCAATCAAT	1261		
DB	1138	-CACAAAGCATCTCTCAAAACATTAATACTCAGSCACATCTGTATCTTTTCA-----	1191		
QY	1262	AGAACATGTTTTTATGTTTGTACAAAAGAAATGTTTATGTTGGTGGGATGAGGTATAG	1321		
DB	1192	-----CTGTCGCTGCTGTTTTCATTTGATGAAAGGCTTGGGTGGGATTTGA	1245		
QY	1322	ACCATGATGTCACCTTCAAGCTACTTAAATAAAGGATCTTAAATAAAGGAGGACT	1381		
DB	1246	AGATGTCAGATCGCTCTGGGTGATTTCAATAAAGGATCTTAAATAAAGGAGGACT	1305		
QY	1382	GTGAACAAGACACCCCTAATAATGTTGATGTTGAAGTGTGAAGTGTGAAGTGTGAAG	1441		
DB	1306	ACGAAGAATCTCTGAAATGAGTTTACGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG	1365		
QY	1442	AACTCTTTT	1450		
DB	1366	GACTCTTTT	1374		
RESULT 6		597 bp mRNA linear SYN 19-AUG-2002			
AF529828		Mus musculus clone 1 transgenic Homo sapiens AID (AID) mRNA,			
LOCUS		complete cds.			
DEFINITION		AF529828			
ACCESSION		AF529828.1 GI:22297243			
VERSION		house mouse.			
KEYWORDS		Mus musculus			
SOURCE		Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.			
ORGANISM		Martin,A. and Scharff,M.D.			
REFERENCE		Somatic hypermutation of the AID transgene in B and non-B cells			
AUTHORS		1 (bases 1 to 597)			
TITLE		Martin,A. and Scharff,M.D.			
JOURNAL		2 (bases 1 to 597)			
REFERENCE		Direct Submision			
AUTHORS		Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of			
TITLE		Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA			
JOURNAL		Location/Qualifiers			
FEATURES		1..597			
source		/organism="Mus musculus"			
		/db_xref="taxon:10090"			
		/clone="1"			
		/cell_type="hybridoma pl-5"			
		/transgenic			

```
source      1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
gene         1..597
            /gene="AID"
            /gene="AID"
            /gene="AID"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95415.1"
            /db_xref="GI:22297244"
            /translation="MDSLLMNRKFLYOFKNRWAKGRRETYLCYVVKRDSATSFSL
            DFGYLRNKNCHVELLFLRYISDWDLDPGRCYRVWFTSWSPCYDCARHVADFLRGNP
            NLSLRIFETARLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFYCWNTFVENHERTFK
            AWGLHENSVRJSRLRILLPLYEVDLRDAFTLGL"
BASE COUNT  129 a 164 c 154 g 150 t
ORIGIN
Query Match      21.1%; Score 595.4; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 4.1e-107;
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTACCAATTCAAAAATGTCGCTGG 139
      |||||
Db 1 ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 199
      |||||
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 259
      |||||
Db 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY 260 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGCGCTTACCGCGTCACCTGG 319
      |||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGCGCTTACCGCGTCACCTGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTACGACTGTGCGCGGCTGCAATAGCCATATGCAATGCAATG 379
      |||||
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCGCGGCTGCAATAGCCATATGCAATGCAATG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTTACTTCTGTGAGGACCGCAAG 439
      |||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGCTGCGGGCGCTGCACCGCGCGCTGCAATAGCCATATGCAATGCAATG 499
      |||||
Db 361 GCTGAGCCGAGGGCTGCGGGCGCTGCACCGCGCGCTGCAATAGCCATATGCAATGCAATG 240
QY 500 TTCAAGATTTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTCAAA 559
      |||||
Db 421 TTCAAGATTTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCTT 619
      |||||
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTGA 597
      |||||
Db 541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTGA 597

RESULT 7
AF529815
LOCUS      Homo sapiens clone Ramos 1 AID (AID) mRNA, partial cds.
DEFINITION
ACCESSION AF529815
VERSION   AF529815.1
KEYWORDS  human.
SOURCE    Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS    Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
           cells
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 596)
AUTHORS    Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL    Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
           Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES   Location/Qualifiers
           1..596
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 1"
            /gene="AID"
            /gene="AID"
            /note="integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95402.1"
            /db_xref="GI:22297218"
            /translation="MDSLLMNRKFLYOFKNRWAKGRRETYLCYVVKRDSATSFSL
            DFGYLRNKNCHVELLFLRYISDWDLDPGRCYRVWFTSWSPCYDCARHVADFLRGNP
            NLSLRIFETARLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFYCWNTFVENHERTFK
            AWGLHENSVRJSRLRILLPLYEVDLRDAFTLGL"
BASE COUNT  128 a 163 c 155 g 150 t
ORIGIN
Query Match      21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTACCAATTCAAAAATGTCGCTGG 139
      |||||
Db 1 ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 199
      |||||
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 259
      |||||
Db 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY 260 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGCGCTTACCGCGTCACCTGG 319
      |||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGCGCTTACCGCGTCACCTGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTACGACTGTGCGCGGCTGCAATAGCCATATGCAATGCAATG 379
      |||||
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCGCGGCTGCAATAGCCATATGCAATGCAATG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTTACTTCTGTGAGGACCGCAAG 439
      |||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGCTGCGGGCGCTGCACCGCGCGCTGCAATAGCCATATGCAATGCAATG 499
      |||||
Db 361 GCTGAGCCGAGGGCTGCGGGCGCTGCACCGCGCGCTGCAATAGCCATATGCAATGCAATG 240
QY 500 TTCAAGATTTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTCAAA 559
      |||||
Db 421 TTCAAGATTTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCTT 619
      |||||
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTG 675
```

```
|||||
Db 541 TTGCCCCCTGATGAGTTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 8
AF529816
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.
ACCESSION AF529816
VERSION AF529816.1 GI:22297219
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 2"
1..>596
/gene="AID"
1..>596
/gene="AID"
/feature="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95403.1"
/db_xref="GI:22297220"
/translation="MDSLNNRRKFLYQFKNVKRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRIYISDWLDPGRCYRVTFWTSKPCYDCARRHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYCWNTFVENHERTFK
AWEGLHENSVRSLRQLRILLPLLYEVDLDRDAFTLGL"
BASE COUNT 127 a 165 c 155 g 149 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 120
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCCACGTGGAATTCGTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCCACGTGGAATTCGTC 180
QY 260 TTCTCCCGCTACATCTCGGACCTGGGACCTAGACCTGGCGCGTGTACCGCGTCACTCGG 319
Db 181 TTCTCCCGCTACATCTCGGACCTGGGACCTAGACCTGGCGCGTGTACCGCGTCACTCGG 240
QY 320 TTCACCTCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTTGGCAGG 379
Db 241 TTCACCTCTCCGAGCCCTCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTTGGCAGG 300
QY 380 AACCCCAACCTCAGTGTAGGATCTTTCACCGCGCGCTCTACTCTGTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTGTAGGATCTTTCACCGCGCGCTCTACTCTGTGTGAGGACCGCAAG 360
|||||

Db 440 GCTGAGCCCGAGGGCTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
|||||
Db 361 GCTGAGCCCGAGGGCTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
|||||
QY 500 TTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 559
|||||
Db 421 TTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 480
|||||
QY 560 GCTCGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
|||||
Db 481 GCTCGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
QY 620 TTGCCCCCTGATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTTTG 675
|||||
Db 541 TTGCCCCCTGATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 9
AF529819
LOCUS Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.
ACCESSION AF529819
VERSION AF529819.1 GI:22297225
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 5"
1..>596
/gene="AID"
1..>596
/gene="AID"
/feature="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95406.1"
/db_xref="GI:22297226"
/translation="MDSLNNRRKFLYQFKNVKRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRIYISDWLDPGRCYRVTFWTSKPCYDCARRHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYCWNTFVENHERTFK
AWEGLHENSVRSLRQLRILLPLLYEVDLDRDAFTLGL"
BASE COUNT 128 a 164 c 154 g 150 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 120
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCCACGTGGAATTCGTC 259
|||||
```

```
Db 121 TCCCTTTTCTGAGGCTTTGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGCTC 180
Qy 260 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 319
Db 181 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 240
Qy 320 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 379
Db 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 300
Qy 380 AACCCCAACTCTAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACTCTAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCGAGGGCTGCGGGGCTGCACCGCGCGGCTGCACCGCGCGGCTGCAATAGCCATGACC 499
Db 361 GCTGAGCCGAGGGCTGCGGGGCTGCACCGCGCGGCTGCACCGCGCGGCTGCAATAGCCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 480
Qy 560 GCTGGAGAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGACTTTG 675
Db 481 GCTGGAGAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGACTTTG 540
Qy 620 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTG 596
Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTG 596

RESULT 10
AF529820
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished
2 (bases 1 to 596)
Martin, A. and Scharff, M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 6"
1. .>596
/gene="AID"
1. .>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95407.1"
/db_xref="GI:22297228"
/translation="NDSLMMRRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELFLRYISDWLDPGRYRVWFTSWSPCYDCARHADVFLRGNP
NLSLRTFLRYFCVDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHNSVRLQLRLLPLLYEVDLDRAFRTGL"
126 a 164 c 155 g 151 t

BASE COUNT
ORIGIN

Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGACACGCTCTTGATGAAACCGGAGGAAGTTCTTTACCAATTTCAAAATTTGTCGCTGG 139
Db 1 ATGACACGCTCTTGATGAAACCGGAGGAAGTTCTTTACCAATTTCAAAATTTGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGAGACCTACCTGCTGCTAGTGTAGAGAGGCGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGGCTGAGACCTACCTGCTGCTAGTGTAGAGAGGCGTGACAGTGTCTACA 120
Qy 200 TCCCTTTTCACTGGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGCTC 259
Db 121 TCCCTTTTCACTGGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGCTC 180
Qy 260 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 319
Db 181 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 240
Qy 320 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 379
Db 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 300
Qy 380 AACCCCAACTCTAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACTCTAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCGAGGGCTGCGGGGCTGCACCGCGCGGCTGCACCGCGCGGCTGCAATAGCCATGACC 499
Db 361 GCTGAGCCGAGGGCTGCGGGGCTGCACCGCGCGGCTGCACCGCGCGGCTGCAATAGCCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 480
Qy 560 GCTGGAGAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGACTTTG 619
Db 481 GCTGGAGAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGACTTTG 540
Qy 620 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCGTTACTTTGGGACTTTG 596

RESULT 11
AF529821
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished
2 (bases 1 to 596)
Martin, A. and Scharff, M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 7"
1. .>596
/gene="AID"
1. .>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95407.1"
/db_xref="GI:22297228"
/translation="NDSLMMRRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELFLRYISDWLDPGRYRVWFTSWSPCYDCARHADVFLRGNP
NLSLRTFLRYFCVDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHNSVRLQLRLLPLLYEVDLDRAFRTGL"
126 a 164 c 155 g 151 t

BASE COUNT
ORIGIN
```


[illegible]

```
RESULT 13
AF529823
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.
ACCESSION AF529823
VERSION AF529823.1 GI:22297233
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 9"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95410.1"
/db_xref="GI:22297234"
/translation="MDSLNNRRKFLYQKKNVWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGHVELFLRYISDWLDPGRCYRVWTFWSWSPCYDCARHVAFLRGNP
NLSLRFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHNSVRLSRQLRILLPLYEVDLDRDAFRTGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGCACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 139
DB 1 ATGCACAGCCTATTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 60
QY 140 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGCGTGACAGTGTCTACA 199
DB 61 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGCGTGACAGTGTCTACA 120
QY 200 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCACGTGGAATTGCTC 259
DB 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCACGTGGAATTGCTC 180
QY 260 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTGCACCTGG 319
DB 181 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTGCACCTGG 240
QY 320 TTCACTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 379
DB 241 TTCACTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 300
QY 380 AACCCCAACTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439
DB 301 AACCCCAACTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGGTGGCGGGGTGTCACCGCGCGGGGTGCAATAGCCATCATGACC 499
|||||
```

```
Db 361 GCTGAGCCCGAGGGCTGCGGGCTGCACCGCGCGGGTGCAAAATAGCATCATGACC 420
QY 500 TTCAAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 559
|||||
Db 421 TTCAAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
|||||
QY 560 GCCTGGGAAGGGCTGCATGAAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
|||||
Db 481 GCCTGGGAAGGGCTGCATGAAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
QY 620 TTGCCCTCTGATGAGGTTGATGACTTACGAGAGCGCAATTCGTAATTTGGGACTTTG 675
|||||
Db 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGAGCGCAATTCGTAATTTGGGACTTTG 596
|||||
RESULT 14
AF529824
LOCUS Homo sapiens clone Ramos 10 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds.
ACCESSION AF529824
VERSION AF529824.1 GI:22297235
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 10"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translation="MDSLNNRRKFLYQKKNVWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKNHVELFLRYISDWLDPGRCYRVWTFWSWSPCYDCARHVAFLRGNP
NLSLRFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHNSVRLSRQLRILLPLYEVDLDRDAFRTGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGCACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 139
DB 1 ATGCACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 60
QY 140 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGCGTGACAGTGTCTACA 199
DB 61 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGCGTGACAGTGTCTACA 120
QY 200 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCACGTGGAATTGCTC 259
DB 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCACGTGGAATTGCTC 180
```

Matches	595;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	80	ATGCACAGCCTCTTGATGAACCGGAGAACTTCTTTACCAATTCAAAAATGCCGCTGG	139						
Db	1	ATGCACAGCCTCTTGATGAACCGGAGAACTTCTTTACCAATTCAAAAATGCCGCTGG	60						
Qy	140	GCTAAGGTCGGCGTGAGACCTACCTGTGTCTAGTGAAGAGGCGTGACAGTGCTACA	199						
Db	61	GCTAAGGTCGGCGTGAGGCTACCTGTGTCTAGTGAAGAGGCGTGACAGTGCTACA	120						
Qy	200	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTCGCTC	259						
Db	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTCGCTC	180						
Qy	260	TTCTCTCGCTACATCTCGGACTCGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACCTGG	319						
Db	181	TTCTCTCGCTACATCTCGGACTCGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACCTGG	240						
Qy	320	TTCACTCTCTGGAGCCCCCTGCTACGACTGTGCCGACATGTGGCCGACACTTTCGCGAGGG	379						
Db	241	TTCACTCTCTGGAGCCCCCTGCTACGACTGTGCCGACATGTGGCCGACACTTTCGCGAGGG	300						
Qy	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	439						
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360						
Qy	440	GCTGAGCCCCGAGGGGCTGCGGGCGGTGCACCCGCGCGGGTGCAAATAGCCATCATGACC	499						
Db	361	GCTGAGCCCCGAGGGGCTGCGGGCGGTGCACCCGCGCGGGTGCAAATAGCCATCATGACC	420						
Qy	500	TTCAAGATTTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA	559						
Db	421	TTCAAGATTTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA	480						
Qy	560	GCTTGGGAAGGGCTGCATGAAAAATTCAGTTCGCTCTCTCCACAGACGTTTCGGCGCATCCTT	619						
Db	481	GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGCTCTCTCCACAGACGTTTCGGCGCATCCTT	540						
Qy	620	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTTTGGGACATTTG	675						
Db	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTTTGGGACATTTG	596						

JOURNAL Unpublished

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:33:05 ; Search time 1524.43 Seconds
(without alignments)
11397.322 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676
Perfect score: 597
Sequence: 1 atggacagcctcttgatgaa.....ttcgacttgggactttga 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

Result No.	Score	Query Match %	Length	DB ID	Description
1	597	100.0	1837	9	BC006296 Homo sapi
2	597	100.0	1791	9	AB040431 Homo sapi
3	595.4	99.7	597	12	AF529828 Mus muscu
4	594.4	99.6	596	9	AF529815 Homo sapi
5	594.4	99.6	596	9	AF529816 Homo sapi
6	594.4	99.6	596	9	AF529819 Homo sapi
7	594.4	99.6	596	9	AF529820 Homo sapi
8	594.4	99.6	596	9	AF529821 Homo sapi
9	594.4	99.6	596	9	AF529822 Homo sapi
10	594.4	99.6	596	9	AF529823 Homo sapi
11	594.4	99.6	596	9	AF529824 Homo sapi
12	594.4	99.6	596	9	AF529825 Homo sapi
13	594.4	99.6	596	9	AF529826 Homo sapi
14	594.4	99.6	596	9	AF529827 Homo sapi
15	594.4	99.6	596	12	AF529829 Mus muscu
16	594.4	99.6	596	12	AF529830 Mus muscu
17	594.4	99.6	596	12	AF529831 Mus muscu
18	594.4	99.6	596	12	AF529833 Mus muscu
19	594.4	99.6	596	12	AF529835 Mus muscu
20	594.4	99.6	596	12	AF529837 Mus muscu
21	594.4	99.6	596	12	AF529839 Mus muscu
22	593.4	99.4	595	12	AF529834 Mus muscu
23	592.8	99.3	596	9	AF529817 Homo sapi
24	592.8	99.3	596	9	AF529818 Homo sapi
25	592.8	99.3	596	12	AF529832 Mus muscu
26	592.8	99.3	596	12	AF529836 Mus muscu
27	592.8	99.3	596	12	AF529838 Mus muscu
28	592.2	99.2	597	12	AF529841 Cricetulu
29	592.2	99.2	597	12	AF529842 Cricetulu
30	592.2	99.2	597	12	AF529843 Cricetulu
31	592.2	99.2	597	12	AF529844 Cricetulu
32	592.2	99.2	597	12	AF529845 Cricetulu
33	592.2	99.2	597	12	AF529846 Cricetulu
34	592.2	99.2	597	12	AF529847 Cricetulu
35	592.2	99.2	597	12	AF529848 Cricetulu
36	592.2	99.2	597	12	AF529849 Cricetulu
37	592.2	99.2	597	12	AF529850 Cricetulu
38	592.2	99.2	597	12	AF529851 Cricetulu
39	592.2	99.2	597	12	AF529852 Cricetulu
40	592.2	99.2	597	12	AF529853 Cricetulu
41	592.2	99.2	597	12	AF529854 Cricetulu
42	586.2	98.2	591	12	AF529855 Cricetulu
43	575.4	96.4	577	12	AF529840 Mus muscu
44	542.2	90.8	547	12	AF529856 Cricetulu
45	475.4	79.6	2440	10	AF132979 Mus muscu

ALIGNMENTS

RESULT 1
BC006296
LOCUS BC006296 1837 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone
MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION BC006296
VERSION MGC.
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1837)
AUTHORS Strausberg, R.
TITLE Direct Submission

Pred. No. is the number of results predicted by chance to have a

JOURNAL

Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapps-remail.nih.gov

Tissue Procurement: Louis Staudt

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hri.nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamen, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.I., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 17 Row: a Column: 1

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9988409.

FEATURES

source

1. .1837
 /organism="Homo sapiens"
 /db_xref="LocusID:57379"
 /db_xref="taxon:9606"
 /clone="MGC:12911 IMAGE:4054915"
 /tissue_type="Primary B-Cells from Tonsils"
 /clone_lib="NIH-MGC_48"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 86. .682
 /codon_start=1
 /product="activation-induced cytidine deaminase"
 /protein_id="AAH06296.1"
 /db_xref="GI:13623401"

CDS

1 ATGGACAGCCTCTTGATGACCGGAGGAAGTTCTTTTACCAATTCAAAATGFCGCTGG 60
 |||||
 86 ATGGACAGCCTCTTGATGACCGGAGGAAGTTCTTTTACCAATTCAAAATGFCGCTGG 145
 |||||
 61 GCTAAGGTCGGCGTCAGACCTACCTGCTAGTAGTAGAGAGGCGTGACAGTGCTACA 120
 |||||
 146 GCTAAGGTCGGCGTCAGACCTACCTGCTAGTAGTAGAGAGGCGTGACAGTGCTACA 205
 |||||
 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATTAAGAAGCGCTGCCACGTGGAATTCGCTC 180
 |||||
 206 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATTAAGAAGCGCTGCCACGTGGAATTCGCTC 265
 |||||
 181 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 240
 |||||
 266 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 325
 |||||
 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACATTTCTCGCCAGGG 300
 |||||
 326 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACATTTCTCGCCAGGG 385
 |||||

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 597; DB 9; Length 1837;
 Best Local Similarity 100.0%; Pred. No. 3.1e-150;
 Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACAGCCTCTTGATGACCGGAGGAAGTTCTTTTACCAATTCAAAATGFCGCTGG 60
 |||||
 Db 86 ATGGACAGCCTCTTGATGACCGGAGGAAGTTCTTTTACCAATTCAAAATGFCGCTGG 145
 |||||
 QY 61 GCTAAGGTCGGCGTCAGACCTACCTGCTAGTAGTAGAGAGGCGTGACAGTGCTACA 120
 |||||
 Db 146 GCTAAGGTCGGCGTCAGACCTACCTGCTAGTAGTAGAGAGGCGTGACAGTGCTACA 205
 |||||
 QY 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATTAAGAAGCGCTGCCACGTGGAATTCGCTC 180
 |||||
 Db 206 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATTAAGAAGCGCTGCCACGTGGAATTCGCTC 265
 |||||
 QY 181 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 240
 |||||
 Db 266 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 325
 |||||
 QY 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACATTTCTCGCCAGGG 300
 |||||
 Db 326 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACATTTCTCGCCAGGG 385
 |||||

QY 301 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
 |||||

Db 386 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 445
 |||||

QY 361 GCTGAGCCCGAGGGCTGCGGGGTGCACCGCGCGGGGTGCAATATGCCATCATGACC 420
 |||||

Db 446 GCTGAGCCCGAGGGCTGCGGGGTGCACCGCGCGGGGTGCAATATGCCATCATGACC 505
 |||||

QY 421 TTCAAGATATTTTACTGCTGGATACCTTTGTAGAAACCATCAAGAACTTTTCAAA 480
 |||||

Db 506 TTCAAGATATTTTACTGCTGGATACCTTTGTAGAAACCATCAAGAACTTTTCAAA 565
 |||||

QY 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGGCGCATCCTT 540
 |||||

Db 566 GCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGGCGCATCCTT 625
 |||||

QY 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTAATTTGGACTTTGA 597
 |||||

Db 626 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTAATTTGGACTTTGA 682
 |||||

RESULT 2

AB040431

LOCUS

DEFINITION

Homo sapiens AID mRNA for activation-induced cytidine deaminase,
 complete CDS.

ACCESSION

AB040431

VERSION

AB040431.1 GI:9988409

KEYWORDS

AID; activation-induced cytidine deaminase; Human AID.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (sites)

AUTHORS

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.

TITLE

Isolation, tissue distribution, and chromosomal localization of the
 human activation-induced cytidine deaminase (AID) gene

JOURNAL

Genomics 68 (1), 85-88 (2000)

MEDLINE

20408890

REFERENCE

2 (sites)

AUTHORS

Rev, P., Muto, T., Levy, Y., Gelssmann, F., Plebani, A., Sanal, O.,
 Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,
 Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N.,
 Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.
 and Durandy, A.

TITLE

Activation-induced cytidine deaminase (AID) deficiency causes the
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)

JOURNAL

Cell 102 (5), 565-575 (2000)

MEDLINE

20460541

REFERENCE

3 (bases 1 to 2791)

AUTHORS

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.

TITLE

Direct Submission

JOURNAL

Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department
 of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,
 Kyoto, Kyoto 606-8501, Japan (E-mail: honjoemfour.med.kyoto-u.ac.jp,
 Tel: 81-75-753-4371(ex.4371), Fax: 81-75-753-4388)

FEATURES

source

1. .2791

/organism="Homo sapiens"

/db_xref="taxon:9606"

1. .2791

/gene="AID"

77. .673

/gene="AID"

/codon_start=1

/product="activation-induced cytidine deaminase"

/protein_id="BAB12721.1"

/db_xref="GI:9988410"

/translation="MDSLNNRRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
 DFGYLRNKGCHVEILLFLRYISDWLDPCRCYRVMTFTSWSPCYDCARHVAQFLRGNP
 NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTVENHRTFK
 AWEGLHENSVRSLRRLILLPLYEVDLDRDAFTLGL"

BASE COUNT	842 a	548 c	625 g	776 t
ORIGIN				
Query Match	100.0%; Score 597; DB 9; Length 2791;			
Best Local Similarity	100.0%; Pred. No. 3.3e-150;			
Matches 597; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG	60	
Db	77	ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG	136	
QY	61	GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTGAAGAGGGGTGACAGTGCTACA	120	
Db	137	GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTGAAGAGGGGTGACAGTGCTACA	196	
QY	121	TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACCTGGAAATGCTC	180	
Db	197	TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACCTGGAAATGCTC	256	
QY	181	TTCCCTCCGTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG	240	
Db	257	TTCCCTCCGTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG	316	
QY	241	TTACCTCTCTGAGCCCTGTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG	300	
Db	317	TTACCTCTCTGAGCCCTGTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG	376	
QY	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360	
Db	377	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	436	
QY	361	GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	420	
Db	437	GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	496	
QY	421	TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAAACATGAAGAACTTTCAAA	480	
Db	497	TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAAACATGAAGAACTTTCAAA	556	
QY	481	GCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT	540	
Db	557	GCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT	616	
QY	541	TTGCCCTCTATGAGTTTGATGACTTTACGAGACGATTTTCGTTGGGACTTTGA	597	
Db	617	TTGCCCTCTATGAGTTTGATGACTTTACGAGACGATTTTCGTTGGGACTTTGA	673	
RESULT 3				
AF529828	597 bp mRNA linear SYN 19-AUG-2002			
LOCUS	Mus musculus clone 1 transgenic Homo sapiens AID (AID) mRNA,			
DEFINITION	complete cds.			
ACCESSION	AF529828			
VERSION	AF529828.1 GI:22297243			
KEYWORDS	house mouse.			
SOURCE	Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.			
REFERENCE	1 (bases 1 to 597)			
AUTHORS	Martin,A. and Scharff,M.D.			
TITLE	Somatic hypermutation of the AID transgene in B and non-B cells			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 597)			
AUTHORS	Martin,A. and Scharff,M.D.			
TITLE	Direct Submission			
JOURNAL	Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA			
FEATURES	Location/Qualifiers			
source	1. .597			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
QY	1	ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG	60	
Db	77	ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG	136	
QY	61	GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTGAAGAGGGGTGACAGTGCTACA	120	
Db	137	GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTGAAGAGGGGTGACAGTGCTACA	196	
QY	121	TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACCTGGAAATGCTC	180	
Db	197	TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACCTGGAAATGCTC	256	
QY	181	TTCCCTCCGTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG	240	
Db	257	TTCCCTCCGTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG	316	
QY	241	TTACCTCTCTGAGCCCTGTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG	300	
Db	317	TTACCTCTCTGAGCCCTGTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG	376	
QY	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360	
Db	377	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	436	
QY	361	GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	420	
Db	437	GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	496	
QY	421	TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAAACATGAAGAACTTTCAAA	480	
Db	497	TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAAACATGAAGAACTTTCAAA	556	
QY	481	GCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT	540	
Db	557	GCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT	616	
QY	541	TTGCCCTCTATGAGTTTGATGACTTTACGAGACGATTTTCGTTGGGACTTTGA	597	
Db	617	TTGCCCTCTATGAGTTTGATGACTTTACGAGACGATTTTCGTTGGGACTTTGA	673	
Query Match	99.7%; Score 595.4; DB 12; Length 597;			
Best Local Similarity	99.8%; Pred. No. 7e-150;			
Matches 596; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG	60	
Db	1	ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG	60	
QY	61	GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTGAAGAGGGGTGACAGTGCTACA	120	
Db	61	GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTGAAGAGGGGTGACAGTGCTACA	120	
QY	121	TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACCTGGAAATGCTC	180	
Db	121	TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACCTGGAAATGCTC	180	
QY	181	TTCTCTCCGTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG	240	
Db	181	TTCTCTCCGTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG	240	
QY	241	TTACCTCTCTGAGCCCTGTGCTACGACTGTGCCGACATGTGGCCGACCTTTCTGCGAGGG	300	
Db	241	TTACCTCTCTGAGCCCTGTGCTACGACTGTGCCGACATGTGGCCGACCTTTCTGCGAGGG	300	
QY	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360	
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360	
QY	361	GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	420	
Db	361	GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	420	
QY	421	TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAAACATGAAGAACTTTCAAA	480	
Db	421	TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAAACATGAAGAACTTTCAAA	480	
QY	481	GCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT	540	
Db	481	GCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTT	540	
QY	541	TTGCCCTCTATGAGTTTGATGACTTTACGAGACGATTTTCGTTGGGACTTTGA	597	
Db	541	TTGCCCTCTATGAGTTTGATGACTTTACGAGACGATTTTCGTTGGGACTTTGA	597	
RESULT 4				
AF529815	596 bp mRNA linear PRI 19-AUG-2002			
LOCUS	Homo sapiens clone Ramos 1 AID (AID) mRNA, partial cds.			
DEFINITION	AF529815			
ACCESSION	AF529815			
KEYWORDS	AF529815.1 GI:22297217			

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 596)
JOURNAL Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 1"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AA095402.1"
/db_xref="GI:22297218"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFRLYISDMDLDPGRCYRVTWFTSPDCARHVAADFURGNP
NLSLRIFARLYFCEDRKAEPGLRLRAGVQIAIMTFKDYFCWNTFVENHETERK
AWEGLEHNSVRLSRIRLLPLLEVDLDRDAFRITGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Query Match 99.6%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGACACGCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Db |
1 ATGACACGCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 61 GCTAAGGTCGCGGTGAGACCTACCTGTGTCTAGTGAAGAGCGGTGACAGTGTCTT 120
Db |
61 GCTAAGGTCGCGGTGAGACCTACCTGTGTCTAGTGAAGAGCGGTGACAGTGTCTT 120
QY 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGAATTGCTC 180
Db |
121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGAATTGCTC 180
QY 181 TTCCTCCGCTACATCTCGGACTGGACCTAGACCTTGGCCGCTGCTACCGGTCACCTGG 240
Db |
181 TTCCTCCGCTACATCTCGGACTGGACCTTGGCCGCTGCTACCGGTCACCTGG 240
QY 241 TTCAAGATATTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGACTTTTCAA 480
Db |
421 TTCAAGATATTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGACTTTTCAA 480
QY 481 GCCTGGAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTT 540
|

Db 481 GCCTGGAGGCTGCATGAAATTCAGTTGCTCTCTCAGACACATTCGGCGCATCCTT 540
QY 541 TTGCCCCCTATGAGGTTGATGACTTAGACAGCAGCATTTGCTACTTTGGGACTTTG 596
|
Db 541 TTGCCCCCTATGAGGTTGATGACTTAGACAGCAGCATTTGCTACTTTGGGACTTTG 596
|
RESULT 5
AF529816 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.
DEFINITION AF529816
ACCESSION AF529816.1 GI:22297219
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 596)
JOURNAL Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AA095403.1"
/db_xref="GI:22297220"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFRLYISDMDLDPGRCYRVTWFTSPDCARHVAADFURGNP
NLSLRIFARLYFCEDRKAEPGLRLRAGVQIAIMTFKDYFCWNTFVENHETERK
AWEGLEHNSVRLSRIRLLPLLEVDLDRDAFRITGL"
BASE COUNT 127 a 165 c 155 g 149 t
ORIGIN

Query Match 99.6%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGACACGCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Db |
1 ATGACACGCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 61 GCTAAGGTCGCGGTGAGACCTACCTGTGTCTAGTGAAGAGCGGTGACAGTGTCTT 120
Db |
61 GCTAAGGTCGCGGTGAGACCTACCTGTGTCTAGTGAAGAGCGGTGACAGTGTCTT 120
QY 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGAATTGCTC 180
Db |
121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGAATTGCTC 180
QY 181 TTCCTCCGCTACATCTCGGACTGGACCTAGACCTTGGCCGCTGCTACCGGTCACCTGG 240
Db |
181 TTCCTCCGCTACATCTCGGACTGGACCTTGGCCGCTGCTACCGGTCACCTGG 240
QY 241 TTCAAGATATTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGACTTTTCAA 300
Db |
241 TTCCTCCGCGAGCGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGG 300
|

BASE COUNT		126 a	164 c	155 g	151 t	/db_xref="taxon:9606"	
ORIGIN						/clone="Ramos 7"	
Query Match		99.6%	Score 594.4;	DB 9;	Length 596;	/gene="AID"	
Best Local Similarity		99.8%;	Pred. No. 1.3e-149;			/gene="AID"	
Matches 595;		Conservative	0;	Mismatches	1;	Indels	0;
Gaps		0;					
Qy	1	ATGACAGCCTCTTGATGAACCGGAGAACTTCTTTACCAATTTCAAAATGTCGCTGG	60				
Db	1	ATGACAGCCTCTTGATGAACCGGAGAACTTCTTTACCAATTTCAAAATGTCGCTGG	60				
Qy	61	GCTAAGGGTCGGGCTGAGACCTACCTGTGCTAGTGAAGAGCGGTGACAGTGGCTC	120				
Db	61	GCTAAGGGTCGGGCTGAGACCTACCTGTGCTAGTGAAGAGCGGTGACAGTGGCTC	120				
Qy	121	TCCTTTTCACTGGACTTGGTTATCTTCGCAATTAAGAAGCGGTGACAGTGGCTC	180				
Db	121	TCCTTTTCACTGGACTTGGTTATCTTCGCAATTAAGAAGCGGTGACAGTGGCTC	180				
Qy	181	TTCTCCGGCTACATCTCGGACTGGGACCTAGACCTTTCACCAATTTCAAAATGTCGCTG	240				
Db	181	TTCTCCGGCTACATCTCGGACTGGGACCTAGACCTTTCACCAATTTCAAAATGTCGCTG	240				
Qy	241	TTCAAGGATTTATTTTACTGCTGGAATTTGTAGAAACCATGAAAGAACTTTCAA	300				
Db	241	TTCAAGGATTTATTTTACTGCTGGAATTTGTAGAAACCATGAAAGAACTTTCAA	300				
Qy	481	GCCTGGGAAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGGCGATCCTT	540				
Db	481	GCCTGGGAAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGGCGATCCTT	540				
Qy	541	TTGCCCTGTATGAGGTTGATGACTTACGAGCGCATTTTCGTA	596				
Db	541	TTGCCCTGTATGAGGTTGATGACTTACGAGCGCATTTTCGTA	596				
RESULT 8							
AF529821							
LOCUS		596 bp mRNA linear PRI 19-AUG-2002					
DEFINITION		Homo sapiens clone Ramos 7 AID (AID) mRNA, partial cds.					
ACCESSION		AF529821					
VERSION		AF529821.1 GI:22297229					
KEYWORDS		human.					
SOURCE		Homo sapiens					
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE		2 (bases 1 to 596)					
AUTHORS		Martin, A. and Scharff, M.D.					
TITLE		Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA					
FEATURES		Location/Qualifiers					
source		1..596					
		/organism="Homo sapiens"					

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES Location/Qualifiers
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 8"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95409.1"
/db_xref="GI:22297232"
/translation="MDSLNNRRKFLYOFKNVRWAKGRRETYLCYVVKRRDTSFSL
DFGLRNKNGCHVELLFRLYISDWLDPCGRYRVTFWTSWSPCYDCARHVADELRGNP
NLSLRFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEK
AWEGLHENSVRLSRQLRILLPLYEVDLLRDAFRTGL"
BASE COUNT 126 a 165 c 155 g 150 t
ORIGIN
Query Match 99.6%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
QY 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGGGTGACAGTGCTACA 120
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGGGTGACCGTGCTACA 120
QY 121 TCCTTTTCACTGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTGCTC 180
Db 121 TCCTTTTCACTGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTGCTC 180
QY 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTCGCCGACTTTCGCGAGGG 300
Db 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTCGCCGACTTTCGCGAGGG 300
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTTACTTCTGTGAGGACCGCAAG 360
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTTACTTCTGTGAGGACCGCAAG 360
QY 361 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACC 420
Db 361 GCTGAGCCCGAGGGGCTGGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACC 420
QY 421 TTCAAAGATATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
Db 421 TTCAAAGATATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
QY 481 GCTGAGGAGGGCTGATGAAATTCAGTTCGCTCTCCAGACAGCTTGGCGCATCTT 540
Db 481 GCTGAGGAGGGCTGATGAAATTCAGTTCGCTCTCCAGACAGCTTGGCGCATCTT 540
QY 541 TTGCCCTCTGATGAGCTTGATGACTTACGAGAGCGCATTTCTGTTTGGGACTTTG 596

Db 541 TTGCCCTCTGATGAGCTTGATGACTTACGAGACGCATTTGCTACTTTGGGACTTTG 596
RESULT 10
AF529823 Homo sapiens clone Ramos 9 AID (AID) mRNA linear PRI 19-AUG-2002
LOCUS
DEFINITION
AF529823 Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.
ACCESSION
AF529823.1 GI:22297233
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES Location/Qualifiers
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 9"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95410.1"
/db_xref="GI:22297234"
/translation="MDSLNNRRKFLYOFKNVRWAKGRRETYLCYVVKRRDTSFSL
DFGLRNKNGCHVELLFRLYISDWLDPCGRYRVTFWTSWSPCYDCARHVADELRGNP
NLSLRFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEK
AWEGLHENSVRLSRQLRILLPLYEVDLLRDAFRTGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Query Match 99.6%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
QY 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGGGTGACAGTGCTACA 120
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGGGTGACAGTGCTACA 120
QY 121 TCCTTTTCACTGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTGCTC 180
Db 121 TCCTTTTCACTGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTGCTC 180
QY 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTCGCCGACTTTCGCGAGGG 300
Db 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTCGCCGACTTTCGCGAGGG 300
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTTACTTCTGTGAGGACCGCAAG 360
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTTACTTCTGTGAGGACCGCAAG 360

```

Qy 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
|||||
Db 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
|||||
Qy 421 TTCAAGATTTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Db 421 TTCAAGATTTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Qy 481 GCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCCTT 540
|||||
Db 481 GCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCCTT 540
|||||
Qy 541 TTGCCCTGTATGAGGTGATGACTTACGAGAGCAGCTTTTCGTACTTTGGGACTTTG 596
|||||
Db 541 TTGCCCTGTATGAGGTGATGACTTACGAGAGCAGCTTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 11
AF529824
LOCUS Homo sapiens clone Ramos 10 AID (AID) mRNA, linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds.
ACCESSION AF529824
VERSION AF529824.1 GI:22297235
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cloned="Ramos 10"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translation="MDSLMNRKLYQFKNVRWAKGRRETVLCYVVKRDSATPSL
DFGYLRNKGHWELFLRYISDWLDPGRCYRVFTWSWPCYDCARHVPFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFK
AWEGHENSVRSLRILLPLYEVDLDRDAFTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN

Query Match 99.6%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCACAGCTCTTGATACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCGTGG 60
|||||
Db 1 ATGCACAGCTCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCGTGG 60
|||||
Qy 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTAGCTAGTAGAAGAGGCGTGACAGTGTACA 120
|||||
Db 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTAGCTAGTAGAAGAGGCGTGACAGTGTACA 120
|||||
Qy 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTTGCTC 180
|||||

```

```

Db 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTTGCTC 180
|||||
Qy 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACCTGG 240
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACCTGG 240
|||||
Qy 241 TTCACCTCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGAGCTTTCTCGGAGGG 300
|||||
Db 241 TTCACCTCTCTGGAGCCCTCTGCTAGGACTGTGCCGACATGTGCCGAGCTTTCTCGGAGGG 300
|||||
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTTCTGTGTGAGGACCGCAAG 360
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTTCTGTGTGAGGACCGCAAG 360
|||||
Qy 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGGGGTGCANATAGCCATCATGACC 420
|||||
Db 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGGGGTGCANATAGCCATCATGACC 420
|||||
Qy 421 TTCAAGATTTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Db 421 TTCAAGATTTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Qy 481 GCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCCTT 540
|||||
Db 481 GCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCCTT 540
|||||
Qy 541 TTGCCCTGTATGAGGTGATGACTTACGAGAGCAGCTTTTCGTACTTTGGGACTTTG 596
|||||
Db 541 TTGCCCTGTATGAGGTGATGACTTACGAGAGCAGCTTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 12
AF529825
LOCUS Homo sapiens clone Ramos 11 AID (AID) mRNA, linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
ACCESSION AF529825
VERSION AF529825.1 GI:22297237
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cloned="Ramos 11"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95412.1"
/db_xref="GI:22297238"
/translation="MDSLMNRKLYQFKNVRWAKGRRETVLCYVVKRDSATPSL
DFGYLRNKGHWELFLRYISDWLDPGRCYRVFTWSWPCYDCARHVPFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFK
AWEGHENSVRSLRILLPLYEVDLDRDAFTLGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN

```


TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 596)

AUTHORS Martin,A. and Scharff,M.D.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES

Location/Qualifiers

source

1. .596

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Ramos 13"

gene

1. .>596

/gene="AID"

CDS

1. .>596

/gene="AID"

/codon_start=1

/note="Integrated into Burkitt's lymphoma cell line Ramos"

/product="AID"

/protein_id="AAM95414.1"

/db_xref="GI:22297242"

translation="MDSLIMNRKFLYQKKNRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFYRISDMDLPGRCYRVTWFTSWSPCYDCARHVDLFRGNP
NLSLRIETARLYFCEDRKAEPGLRLRHAGVOIAIMTFKDYFCWNTFVENHERTEK
AWEGUHLHENSRLSRLRLILLPLLEYVDLDRDAFRTLGL"

BASE COUNT 127 a 163 c 155 g 151 t

ORIGIN

Query Match 99.6%; Score 594.4; DB 9; Length 596;

Best Local Similarity 99.8%; Pred. No. 1.3e-149;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGACGCCCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

Db 1 ATGGACGCCCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

QY 61 GCTAAGGTCGGCTGAGACCTACCTGTGTACGTAGTGTAGAGGCGTGACGTGCTACA 120

Db 61 GCTAAGGTCGGCTGAGACCTACCTGTGTACGTAGTGTAGAGGCGTGACGTGCTACA 120

QY 121 TCCTTTTCACTGGACTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAAATTCCTC 180

Db 121 TCCTTTTCACTGGACTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAAATTCCTC 180

QY 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCTTACCTACCAATTCAAAAATGTCGGCTGG 240

Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCTTACCAATTCAAAAATGTCGGCTGG 240

QY 361 GCTGAGCCGAGGGCTGCGGGCGGTGACCGCGCGGTGCAATAGCCATCATGACC 420

Db 361 GCTGAGCCGAGGGCTGCGGGCGGTGACCGCGCGGTGCAATAGCCATCATGACC 420

QY 421 TTCAAAGATATTTTACTGCTGGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTT 540

Db 421 TTCAAAGATATTTTACTGCTGGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTT 540

QY 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCCTACTTTGGGACTTTG 596

Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCCTACTTTGGGACTTTG 596

RESULT 15

AF529829

LOCUS

DEFINITION Mus musculus clone 2 transgenic Homo sapiens AID (AID) mRNA, complete cds.

ACCESSION AF529829

VERSION AF529829.1

KEYWORDS GI:22297245

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 596)

Martin,A. and Scharff,M.D.

Somatic hypermutation of the AID transgene in B and non-B cells

Unpublished

2 (bases 1 to 596)

Martin,A. and Scharff,M.D.

Direct Submission

Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES

Location/Qualifiers

source

1. .596

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="2"

/cell_type="hybridoma pl-5"

transgenic

1. .596

/organism="Homo sapiens"

/db_xref="taxon:9606"

gene

1. .>596

/gene="AID"

CDS

1. .>596

/gene="AID"

/codon_start=1

/product="AID"

/protein_id="AAM95416.1"

/db_xref="GI:22297246"

translation="NDRLLMNRKFLYQKKNRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFYRISDMDLPGRCYRVTWFTSWSPCYDCARHVDLFRGNP
NLSLRIETARLYFCEDRKAEPGLRLRHAGVOIAIMTFKDYFCWNTFVENHERTEK
AWEGUHLHENSRLSRLRLILLPLLEYVDLDRDAFRTLGL"

BASE COUNT 128 a 163 c 155 g 150 t

ORIGIN

Query Match 99.6%; Score 594.4; DB 12; Length 596;

Best Local Similarity 99.8%; Pred. No. 1.3e-149;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGACGCCCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

Db 1 ATGGACGCCCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

QY 61 GCTAAGGTCGGCTGAGACCTACCTGTGCTTACCTAGTGTAGAGGCGTGACGTGCTACA 120

Db 61 GCTAAGGTCGGCTGAGACCTACCTGTGCTTACCTAGTGTAGAGGCGTGACGTGCTACA 120

QY 121 TCCTTTTCACTGGACTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAAATTCCTC 180

Db 121 TCCTTTTCACTGGACTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAAATTCCTC 180

QY 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCTTACCTACCAATTCAAAAATGTCGGCTGG 240

Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCTTACCAATTCAAAAATGTCGGCTGG 240

QY 241 TTCACTTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTCTTCCGAGGG 300

Db 241 TTCACTTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTCTTCCGAGGG 300

QY 301 AACCCCAACCTCAGTCTGAGGATTCACCGCGCGCTTACTTCTGTGAGGACCGCAAG 360

Db 301 AACCCCAACCTCAGTCTGAGGATTCACCGCGCGCTTACTTCTGTGAGGACCGCAAG 360

QY 361 GCTGAGCCGAGGGCTGCGGGCGGTGACCGCGCGGTGCAATAGCCATCATGACC 420

Db 361 GCTGAGCCGAGGGCTGCGGGCGGTGACCGCGCGGTGCAATAGCCATCATGACC 420

QY 421 TTCAAAGATATTTTACTGCTGGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTT 540

Db 421 TTCAAAGATATTTTACTGCTGGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTT 540

QY 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCCTACTTTGGGACTTTG 596

Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTCCTACTTTGGGACTTTG 596

Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCCGGCGCCTCTACTTCTGTGTGAGGACCGCAAG	360
Qy	361	GCTGAGCCCGAGGGGCTGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420
Db	361		
Db	361	GCTGAGCCCGAGGGGCTGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420
Qy	421	TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA	480
Db	421		
Qy	481	GCCTGGGAAGGCTGCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTT	540
Db	481		
Qy	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGATTTCTACTTTGGGACTTTG	596
Db	541		

Search completed: June 19, 2003, 07:46:31
Job time : 1525.43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 20:16:19 ; Search time 122.222 Seconds
(without alignments)
11000.051 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676
Perfect score: 597
Sequence: 1 atgacagcctcttgatgaa.....ttcgtacttgggactttga 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				N_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*						
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*						
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*						
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*						
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*						
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*						
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*						
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*						
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*						
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*						
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*						
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*						
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*						
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*						
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*						
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*						
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*						
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*						
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*						
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*						
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*						
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*						
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*						
24:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	ID	Description			
1	597	100.0	2818	21 AAC55312	Human activation-i		
2	475.4	79.6	2440	21 AAC55307	Mouse activation-i		
3	273.4	45.8	6564	21 AAC55314	Human activation-i		
4	273.4	45.8	11204	21 AAC55339	Human activation-i		
5	271	45.4	271	21 AAC55317	Human activation-i		
6	148	24.8	148	21 AAC55316	Human activation-i		
7	118.4	19.8	1534	20 AA220856	Polynucleotide seq		
8	118.4	19.8	1534	22 AAS59293	Human cDNA encodin		
9	118.4	19.8	1534	22 ABA90962	Human polynucleoti		

10	116.8	19.6	1446	21 AA12409	cDNA encoding a hu
11	116.6	19.5	1055	24 AAD24392	Human RNA metaboli
12	114.8	19.2	610	19 AAV48231	Human RNA editing
13	114.8	19.2	610	21 AA72058	cDNA encoding huma
14	114.8	19.2	950	22 AAS41420	cDNA encoding nove
15	114.8	19.2	987	21 AAF16264	Human prostate can
16	114.8	19.2	987	22 AAS41407	cDNA encoding nove
17	114.8	19.2	1120	22 AAI59847	Human polynucleoti
18	114.8	19.2	1143	22 AAI58061	Human polynucleoti
19	114.4	19.2	116	21 AAC55318	Human activation-i
20	113.4	19.0	944	24 ABL99876	Human secretory po
21	110.4	18.5	1348	24 ABN96785	Gene #3283 used to
22	105.4	17.7	1567	22 AAI93080	Human polynucleoti
23	104.4	17.5	2151	24 AAS62572	cDNA sequence #359
24	103.6	17.4	819	22 AAS41669	cDNA encoding nove
25	90.2	15.1	201143	24 ABK83568	Human DNA differen
26	89.8	15.0	716	20 AAZ15848	Human gene express
27	88.6	14.8	6131	22 AAS42187	Genomic sequence #
28	88.6	14.8	6131	22 AAK69784	Human immune/haema
29	71.6	12.0	2338	22 AAI17654	Human cDNA sequenc
30	67	11.2	572	22 ABA63759	Human foetal liver
31	67	11.2	572	22 ABA30950	Probe #9416 for ge
32	67	11.2	572	22 AAK12273	Human brain expres
33	67	11.2	572	22 AAK37995	Human bone marrow
34	67	11.2	572	22 AAI18764	Probe #8697 for ge
35	67	11.2	572	22 AAI43877	Probe #12563 used
36	67	11.2	572	24 ABS11997	Human genome-deriv
37	63.6	10.7	892	19 AAV17184	cDNA encoding a no
38	62	10.4	675	24 AAD25771	Human APOBEC2 cDNA
39	62	10.4	1164	24 AAD27214	Human AAD45360 pro
40	62	10.4	4812	22 AAK77995	Human immune/haema
41	62	10.4	4812	22 AAK84817	Human immune/haema
42	62	10.4	12600	24 AAD25770	Human APOBEC2 gene
43	62	10.4	12600	24 AAD25817	Human APOBEC2 gene
44	59.2	9.9	454	21 AAC00672	Human secreted pro
45	58.4	9.8	879	15 AAQ71633	Apo-B RNA editing

ALIGNMENTS

RESULT 1
AAC55312
ID AAC55312 standard; cDNA; 2818 BP.

XX AAC55312;

DT 05-FEB-2001 (first entry)

Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.

Activation-induced cytidine deaminase; Immune related disease; allergy; allergic disease; antiallergic; antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological; gene therapy; B cell associated immune system disorder; food allergy; immunodeficiency disease; immunoglobulin A deficiency disease; asthma; Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis; drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS; ataxia telangiectasia; common variable immunodeficiency disorder; major histocompatibility class II deficiency disease; auto immunodeficiency syndrome; Igc subclass selection disorder; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
CDS 80..676
/*tag= a
/*product= "activation-induced cytidine deaminase"

WO2000058480-A1.

05-OCT-2000.

XX

PF 28-MAR-2000; 2000WO-JP01918.
 XX 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX
 PA (NISB) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 PA Honjo T, Muramatsu M;
 PI WPI; 2000-611715/58.
 XX P-PSDB; AAB24198.
 DR
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 PS Claim 3; Page 135-139; 174pp; Japanese.
 XX
 CC The present sequence encodes human activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX
 XX Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;
 XX
 Query Match 100.0%; Score 597; DB 21; Length 2818;
 Best Local Similarity 100.0%; Pred. No. 2,9e-167;
 Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGACAGCCTCTGTATGACCGGAGGAGTCTTTTACCAATTCAAAATGTCGGTGG 60
 DB 80 ATGACAGCCTCTGTATGACCGGAGGAGTCTTTTACCAATTCAAAATGTCGGTGG 139
 QY 61 GCTAAGGTCGGCGTGAGACCTACCTGCTAGTGTAGAGGCGTGACATGTGTACA 120
 DB 140 GCTAAGGTCGGCGTGAGACCTACCTGCTAGTGTAGAGGCGTGACATGTGTACA 199
 QY 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATGCTC 180
 DB 200 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATGCTC 259
 QY 181 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCGCTCACTGG 240
 DB 260 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCGCTCACTGG 319
 QY 241 TTCACCTCTCGGACCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGAGGG 300
 DB 320 TTCACCTCTCGGACCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGAGGG 379
 QY 301 AACCCCAACTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
 DB 380 AACCCCAACTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439
 QY 361 GCTGAGCCGAGGGCTGGCGGCTGTACCGCGCGGGGTGCAATAGCCATCATGACC 420
 DB 440 GCTGAGCCGAGGGCTGGCGGCTGTACCGCGCGGGGTGCAATAGCCATCATGACC 499
 QY 421 TTCAAGATATTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGACTTTCAAA 480
 XX

Db 500 TTCAAGATATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 559
 QY 481 GCCTGGGAAGGCTGCATCAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGGCGCATCCTT 540
 Db 560 GCCTGGGAAGGCTGCATCAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGGCGCATCCTT 619
 QY 541 TTGCCCTCTGTATGAGTTGATGACATTACGAGACGCAATTCGTACTTTGGGACTTTGA 597
 Db 620 TTGCCCTCTGTATGAGTTGATGACATTACGAGACGCAATTCGTACTTTGGGACTTTGA 676

RESULT 2

AAC55307

ID AAC55307 standard; cDNA; 2440 BP.

XX AAC55307;

XX 05-FEB-2001 (first entry)

XX Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.
 DE Activation-induced cytidine deaminase; AID; cytidine deaminase;
 XX immune related disease; allergy; allergic disease; antiallergic;
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.

XX Mus musculus.

OS Key Location/Qualifiers

XX 93..689

FT /*tag= a

FT /product= "activation-induced cytidine deaminase"

XX WO200058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

PR 27-DEC-1999; 99JP-0371382.

XX (NISB) JAPAN TOBACCO INC.

PA (HONJ/) HONJO T.

PA Honjo T, Muramatsu M;

PI WPI; 2000-611715/58.

XX P-PSDB; AAB24197.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -

XX Claim 3; Page 126-130; 174pp; Japanese.

XX The present sequence encodes mouse activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen

AAC55339
 ID AAC55339 standard; DNA; 11204 BP.
 XX
 AC AAC55339;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.
 XX
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; antiallergic;
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; atopic disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
 OS Homo sapiens.
 XX
 XX WO200058480-A1.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 28-MAR-2000; 2000WO-JP01918.
 PF
 XX
 XX 29-MAR-1999; 99JP-0087192.
 PR
 XX 24-JUN-1999; 99JP-0178999.
 PR
 XX 27-DEC-1999; 99JP-0371382.
 XX
 PA (NISR) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX
 XX Honjo T, Muramatsu M;
 PI
 XX WPI; 2000-611715/58.
 DR
 XX
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 XX Claim 17; Page 163-170; 174pp; Japanese.
 PS
 XX The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents a genomic DNA sequence of human AID.
 XX
 XX Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;
 SQ
 Query Match 45.8%; Score 273.4; DB 21; Length 11204;
 Best Local Similarity 99.6%; Pred. No. 1.3e-70;
 Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 155 AGAAGCGGTGCCACGTGAATTCCTTCCTCCGCTACATCTCGGACTGGGACCTAGACC 214
 DB 7805 AGAAGCGGTGCCACGTGAATTCCTTCCTCCGCTACATCTCGGACTGGGACCTAGACC 7864

QY 215 CTGGCGCGTGTACCGCGTCACTGGTTCACCTCTCGAGCCCTGCTACGACTGTGCC 274
 DB 7865 CTGGCGCGTGTACCGCGTCACTGGTTCACCTCTCGAGCCCTGCTACGACTGTGCC 7924
 QY 275 GACATGTGGCGGACTTCTCTGCGAGGGAACCCCAACCTCAGTGTGAGGATCTTCACCGCGC 334
 DB 7925 GACATGTGGCGGACTTCTCTGCGAGGGAACCCCAACCTCAGTGTGAGGATCTTCACCGCGC 7984
 QY 335 GCCTTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGGCGGCTGCACCGCG 394
 DB 7985 GCCTTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGGCGGCTGCACCGCG 8044
 QY 395 CGGGGTGCAATAGCCATCATGACCTTCAAGAT 429
 DB 8045 CGGGGTGCAATAGCCATCATGACCTTCAAGAT 8079
 RESULT 5
 AAC55317
 ID AAC55317 standard; DNA; 271 BP.
 XX
 AC AAC55317;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 XX Human activation-induced cytidine deaminase exon 3 SEQ ID NO:13.
 DE
 XX
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; antiallergic;
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; atopic disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
 OS Homo sapiens.
 XX
 XX WO200058480-A1.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 28-MAR-2000; 2000WO-JP01918.
 PF
 XX
 XX 29-MAR-1999; 99JP-0087192.
 PR
 XX 24-JUN-1999; 99JP-0178999.
 PR
 XX 27-DEC-1999; 99JP-0371382.
 XX
 PA (NISR) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX
 XX Honjo T, Muramatsu M;
 PI
 XX WPI; 2000-611715/58.
 DR
 XX
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 XX Claim 18; Page 151; 174pp; Japanese.
 PS
 XX The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen

CC disease, DiGeorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class)
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC encoding AID may be used for gene therapy and the antibodies to the AID
CC protein may be used for diagnosis and treatment of these disorders. The
CC present sequence represents the exon 3 genomic DNA sequence of human AID.
XX
SQ Sequence 271 BP; 47 A; 95 C; 76 G; 53 T; 0 other;

Query Match 45.4%; Score 271; DB 21; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.2e-70;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 157 AACGGCTGCCAGGTGGATTGCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCT 216
DB 1 AACGGCTGCCAGGTGGATTGCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCT 60
QY 217 GGCGCTGCTACCGGCTCACCTGGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGA 276
DB 61 GGCGCTGCTACCGGCTCACCTGGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGA 120
QY 277 CATGTGGCGGACTTCTCGGAGGAAACCCCAACCTCAGTCTGAGGATCTTACCGCGGC 336
DB 121 CATGTGGCGGACTTCTCGGAGGAAACCCCAACCTCAGTCTGAGGATCTTACCGCGGC 180
QY 337 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCGCGGCTGCGGCGCTGCACCGGCC 396
DB 181 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCGCGGCTGCGGCGCTGCACCGGCC 240
QY 397 GGGGTGCAATAGCCATCATGACCTTCAAAG 427
DB 241 GGGGTGCAATAGCCATCATGACCTTCAAAG 271

RESULT 6
AAC55316
ID AAC55316 standard; DNA: 148 BP.
XX
AC AAC55316;
DT 05-FEB-2001 (first entry)
XX
DE Human activation-induced cytidine deaminase exon 2 SEQ ID NO:12.
XX
KW Activation-induced cytidine deaminase; AID: cytidine deaminase;
KW immune related disease; allergy; allergic disease; antiallergic;
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
XX
OS Homo sapiens.
XX
PN WO200058480-A1.
XX
PD 05-OCT-2000.
XX
PF 28-MAR-2000; 2000WO-JP01918.
XX
PR 29-MAR-1999; 99JP-0087192.
PR 24-JUN-1999; 99JP-0178999.
PR 27-DEC-1999; 99JP-0371382.
XX
PA (NISR) JAPAN TOBACCO INC.
PA (HONJ) HONJO T.
XX
PI Honjo T, Muramatsu M; /

DR WPI: 2000-611715/58.
XX
PT Nucleic acid encoding activation induced cytidine deaminase, useful as
PT a target for drug development for immune-related diseases including
PT allergies -
XX
PS Claim 18; Page 150; 174pp; Japanese.
XX

The present invention describes an activation-induced cytidine deaminase
(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
has cytidine activity similar to APOBEC-1. AID has antiallergic,
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
CC dermatological activities, and can be used in gene therapy. AID
CC polynucleotides are useful in methods for identifying drugs for the
CC treatment of B cell associated immune system disorders, immunodeficiency
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
CC disease, DiGeorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class) class
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC encoding AID may be used for gene therapy and the antibodies to the AID
CC protein may be used for diagnosis and treatment of these disorders. The
CC present sequence represents the exon 2 genomic DNA sequence of human AID.
XX
SQ Sequence 148 BP; 35 A; 33 C; 38 G; 42 T; 0 other;

Query Match 24.8%; Score 148; DB 21; Length 148;
Best Local Similarity 100.0%; Pred. No. 4e-34;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 CCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATGTCGCTGGCTAAGGG 68
DB 1 CCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATGTCGCTGGCTAAGGG 60
QY 69 TCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGCTGACAGTGCTACATCTTTTC 128
DB 61 TCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGCTGACAGTGCTACATCTTTTC 120
QY 129 ACTGGACTTGGTTATCTTCGCAATAAG 156
DB 121 ACTGGACTTGGTTATCTTCGCAATAAG 148

RESULT 7
AAZ20856
ID AAZ20856 standard; cDNA: 1534 BP.
XX
AC AAZ20856;
DT 09-DEC-1999 (first entry)
XX
DE Polynucleotide sequence of the lp547_4 clone.
XX
KW secreted protein; cDNA library; clone; transmembrane protein;
KW signal sequence cloning; hybridization cloning; gene therapy;
KW receptor; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 51..1205
FT /tag= a
FT /product= lp547_4
FT sig_peptide 882..980
FT /tag= b
FT mat_peptide 981..1205
FT /tag= c
XX
PN WO9942470-A1.
XX
PD 26-AUG-1999.

XX	18-FEB-1999;	99WO-US03458.	
PF	18-FEB-1999;	98US-0075038.	
XX	17-FEB-1999;	99US-0251600.	
PR	17-FEB-1999;	99US-0251600.	
XX	(GEMV) GENETICS INST INC.		
PA	Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Merberg D;		
XX	Treacy M, Agostino MJ, Steining RJ;		
PI	WPI; 1999-518580/43.		
DR	P-PSDB: AAY42383.		
DR	New polynucleotides encoding human secreted proteins used for		
XX	therapeutic, diagnostic and research purposes.		
PT	Claim 16; Page 104; 125pp; English.		
PT	This is the polynucleotide sequence of the clone lp547.4, which was		
XX	isolated from a human fetal brain cDNA library using methods which are		
CC	selective for cDNAs encoding secreted proteins, or by identification as		
CC	a secreted or transmembrane protein on the basis of computer analysis of		
CC	the amino acid sequence of the encoded protein.		
CC	The PNs and proteins of the invention are predicted to have biological		
CC	activities which would make them suitable for treating, preventing or		
CC	ameliorating medical conditions in humans and animals, although no		
CC	supporting data is given. Suggested activities include nutritional		
CC	activity, cytokine and cell proliferation/differentiation activity,		
CC	immune stimulating (e.g. as vaccines) or suppressing activity,		
CC	hematopoiesis regulating activity, tissue growth activity,		
CC	activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic		
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory		
CC	activity, cadherin/tumor invasion suppressor activity, and tumor		
CC	inhibition activity. The PNs are also stated to be useful for gene		
CC	therapy. Other activities include inhibiting the growth, infection or		
CC	function of bacteria, fungi, viruses and other parasites; effecting		
CC	bodily characteristics such as, e.g. weight, color, skin, etc.,		
CC	effecting biorhythms or circadian cycles; enhancing fertility; treatment		
CC	of depression; treatment of pain; hormonal or endocrine activity.		
XX			
SQ	Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;		
	Query Match	19.8%; Score 118.4; DB 20; Length 1534;	
	Best Local Similarity 58.9%; Pred. No. 7.7e-25;		
	Matches 225; Conservative 0; Mismatches 151; Indels 6; Gaps 1;		
QY	158	ACGGCTGCCACGTGGAAATTCCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTG	217
Db	811	AAGCCGCCATGCAGAGCTGTGCTTCCTGACGATTCCTTTTGGAACTGGACCTGG	870
QY	218	CGCGCTGCTACCCGCGTCACCTGGTTCACCTTCCTGGAGCCCTGCTGACGACTGTCCCGCAC	277
Db	871	ACCAGGACTACAGGGTTACCTGCTTCACCTCTGGAGCCCTGCTTCAGCTGTGCCCAGG	930
QY	278	ATGTGGCGGACTTTCGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCC	337
Db	931	AAATGGCTAAATTCATTTCAAAAAAACAACACGTGAGCCTGTGCATCTTCACCTGCCCGCA	990
QY	338	TCCTACTCTGTGAGGACCGAAGGCTCAGCCCGAGGGGCTGCGCGGCTGCACGCGCCG	397
Db	991	TCCTA-----TGATGATCAAGGAAGATGTCAGGAGGGGCTGCGCACCTGGCGAGGCTG	1044
QY	398	GGGTGCAAAATAGCCATCATGACCTTCAAGATTTATTTTACTGCTGGAATCTTTTCTAG	457
Db	1045	GGGCCAAAATTTCAATAATGACATACAGTGAATTTAAGCACTGCTGGACACCTTTGTGG	1104
QY	458	AAACCATGAAGAATCTTCAAGACCTTGGGAAGGCTGCATGAAAAATTCAGTTGCTGCT	517
Db	1105	ACCACAGGGATGTCCCTTCCAGCCTGGGATGGACTAGATGAGCACAGCAAGACCTGA	1164
QY	518	CCAGACAGCTTCGGCGCATCCT	539

CC contraceptive based on the ability of inhibins to decrease fertility in
CC female mammals and decrease spermatogenesis in male mammals. The
CC proteins and nucleic acids are also useful as food supplements. The
CC present sequence encodes a secreted protein of the invention.

XX Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;

SQ Query Match 19.8%; Score 118.4; DB 22; Length 1534;

Best Local Similarity 58.9%; Pred. No. 7.7e-25;

Matches 225; Conservative 0; Mismatches 151; Indels 6; Gaps 1;

QY 158 ACGGCTGCCAGTGGAAATGCTCTTCCTCCGCTACATCTCGAGTGGACCTAGACCCCTG 217

DB 811 AAGCGCCCATGCAGAGCTGTCTCTCGACGTGATTCCTTTTGGAAAGCTGGACCTGG 870

QY 218 GCCGCTGCTACCGGTCACCTGGTTCACCTCCTCGAGCCCTGTACACACTGTGCCGAC 277

DB 871 ACCAGGACTACAGGGTTACCTGCTTCACCTCCTCGAGCCCTGTCTAGCTGTGCCGAG 930

QY 278 ATGTGGCGACTTTCTTGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCC 337

DB 931 AAATGGCTAAATTCATTTCAAAAACAAACAGCTGAGCCTGTGCATCTTCACTGCCGCA 990

QY 338 TCTACTTCTGTGAGACCGCAAGGCTGAGCCGAGGGCTCGGGGGCTGCACCGCGCG 397

DB 991 TCTA-----TGATGATCAAGGAAGATGTCAGGAGGGCTGCGCACCTCGCGAGGCTG 1044

QY 398 GGGTGCAATAGCCATCATGACCTTCAAGATATTTTACTGCTGGGAATCTTTGTAG 457

DB 1045 GGGCAAAATTTCAATAATGACATACAGTGAATTTAAGCACTGCTGGGACACCTTTTGG 1104

QY 458 AAAACCATGAAGAACTTTTAAAGCTTGGGAGGGCTGCATGAAATTCAGTCTGCTCT 517

DB 1105 ACCACGAGGATGTCCTTCCAGCCCTGGGATGAGTACAGCACAGCCCAAGACCTGA 1164

QY 518 CCAGACAGCTTCGGGGCATCCT 539

DB 1165 GTGGAGGCTGCGGGCCATTCT 1186

RESULT 9

ABA90962

ID ABA90962 standard; cDNA; 1534 BP.

XX ABA90962;

AC ABA90962;

DT 14-FEB-2002 (first entry)

DE Human polynucleotide SEQ ID NO 173.

XX Human; clone bd306-7; ATCC number 98599; gene therapy;

KW immune disorder; bacterial infection; fungal infection; cancer; tumour;

KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;

KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;

KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;

KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerability;

KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;

KW Crohn's disease; cystostatic; anti-inflammatory; immunomodulatory;

KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.

XX Homo sapiens.

OS US2001039335-A1.

XX PN 08-NOV-2001.

PD 04-DEC-2000; 2000US-0729674.

XX 10-APR-1997; 97US-126425P.

PR 04-DEC-1997; 97US-067454P.

PR 20-DEC-1997; 97US-068379P.

PR 02-JAN-1998; 98US-070346P.

PR 07-JAN-1998; 98US-070643P.

PR 08-JAN-1998; 98US-070755P.
PR 13-JAN-1998; 98US-071304P.
PR 22-JAN-1998; 98US-072134P.
PR 30-JAN-1998; 98US-073095P.
PR 18-FEB-1998; 98US-075038P.
PR 30-MAR-2000; 2000US-0539330.
PR 23-NOV-1998; 98US-0197886.
XX (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (AGOS/) AGOSTINO M J.
PA (STEI/) STEININGER R J.
PA (SPAU/) SPAULDING V.
PA (WONG/) WONG G G.
PA (CLAR/) CLARK H.
PA (FECH/) FECHTEL K.

PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark H, Fechtel K;
XX WPI: 2002-040725/05.
DR P-PSDB; ABB55784.
XX New secreted proteins and encoding polynucleotides, useful in gene
therapies, particularly for preventing or treating autoimmune
disorders, cancer, graft-versus-host disease, wound, osteoporosis,
stroke or inflammations

PS Disclosure; Page 310-311; 349pp; English.

XX The invention relates to isolated polynucleotides (ABA90876-ABA90968 and
ABA90980) and encoded proteins (ABB55698-ABB55800), especially
polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and
proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in
clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
are deposited with the American Type Culture Collection (ATCC) with
accession number 98599. The polynucleotides and encoded polypeptides have
cytostatic, anti-inflammatory, immunomodulatory, vulnerary,
neuroprotective, activin, inhibit, chemotactic, haemostatic, thrombolytic
and anti-inflammatory activity and acting as cytokine modulators,
haematopoiesis regulators, tissue growth modulators and/or cadherin
suppressors. The polypeptides and polynucleotides are useful in gene
therapies, particularly for preventing, treating or ameliorating any of
the following diseases: immune deficiency and disorders; e.g. bacterial
or fungal infections, autoimmune disorders, cancer, systemic lupus
erythematosus or graft-versus-host disease; myeloid or lymphoid cell
deficiencies; wound, burns, incisions and ulcers, osteoporosis or
osteoarthritis; central and peripheral nervous system diseases and
neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
or systemic inflammatory response syndrome, ischaemia-reperfusion
injury, endotoxin lethality, arthritis, inflammatory bowel disease or
Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
foliaceus.

SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;

XX Query Match 19.8%; Score 118.4; DB 24; Length 1534;

Best Local Similarity 58.9%; Pred. No. 7.7e-25;

Matches 225; Conservative 0; Mismatches 151; Indels 6; Gaps 1;

QY 158 ACGGCTGCCAGTGGAAATGCTCTTCCTCCGCTACATCTCGAGTGGACCTAGACCCCTG 217

DB 811 AAGCGCCCATGCAGAGCTGTCTTCCTGGAGCTGATTTCCCTTTTGGAAAGCTGGACCTGG 870

QY 218 GCCGCTGCTACCGGCTACCTGCTTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGAC 277

Matches	289;	Conservative	0;	Mismatches	232;	Indels	15;	Gaps	2;
Qy	16	ATGAACCGGAGGAAGTTCTTTTACC	AATTCAAAAATGTCCGGCTGAGGCTAAGGGTCGGCGT	75					
Dd	34	ATGTATCACGACCATTTCTACTTCCA	ATTATAAACCTATGGAAGCAACGATCGGAAC	93					
Qy	76	GAGACCTACCTGTGCTACGTAGTGA	GAGGGCTGACAGTGCTACATCCTTTTTCAC	TCTGCAC	135				
Dd	94	GAACAATTGGCTGTCTTCA	CGGTGGAAGGTATAAAGCGCCGCTCAGTTGTCTCCTGGAA	153					
Qy	136	TTTGGTTATCTTCGCAA-----	TAAGAAGCGCTGCCAGCTGGAATTGGCTCTTC	183					
Dd	154	ACGGCGCTTCCGAACACAGTGGAT	CTTGAGACCCATTGTGACAGAAAGGTGCTTC	213					
Qy	184	CTCGCTACATCTCGGACTGGGACCT	PAGACCTGTGCGCGCTGTACCGGCTCACCTGGTTC	243					
Dd	214	CTCTCTTGGTTCTCGGACGACACTG	CTCTCTTAACACAAAGTACCAAGTCACTTGGTAC	273					
Qy	244	ACCTCCTGGAGCCCCTCTACGACTG	PGCCGACATGTGGCCGACTTTCTGCGAGGGAAAC	303					
Dd	274	ACATCTTGGAGCCCTTGCCACAGACT	GTGCAGGGGAGGTGGCCGAGTTCTGGCCAGGCAC	333					
Qy	304	CCCAACCTCAGTCTGAGGATCTTCA	CAGCGCGCGCTCTACTTCTGTGAGAGCCGCAAGGCT	363					
Dd	334	AGCAACGTGAATCTCACCATCTTCA	CGCCGCTCTACTTCTCAGTATC---CATGT	390					
Qy	364	GAGCCGAGGGGCTGGCGGCTGCAC	CGCGCGGGGTGCAAAATAGCCATCATGACCTTC	423					
Dd	391	TACCAGGAGGGGCTCCGACGCTGAT	GCAGAAAGGGTGCCTGTGAGATCATGACATAT	450					
Qy	424	AAAGATTATTTTTACTGCTGGAATAC	TTTTGTAGAAAAACATGAAGAATTTCAAAGCC	483					
Dd	451	GAAGATTTAAATATTGTGGAAAC	TTTGTGTACAATGATAATGAGCCATTCAAGCT	510					
Qy	484	TGGGAGGGCTGCATGAAATTCNGT	CTGCTCTCCAGACAGCTTCGGCGCATCCT	539					
Dd	511	TGGAAGGATTAAAAACCAACTTTC	GACTTTCGAGTTCTGAAAAGAGGCTACGGGAGAGTCT	566					
RESULT 13									
AAAT72058									
ID	AAAT72058 standard; cDNA; 610 BP.								
XX	AAAT72058;								
XX	24-NOV-2000 (first entry)								
XX	cDNA encoding human RNA editing enzyme REE-2.								
XX	RNA editing enzyme; REE-2; human; HEPR homologue; REPR homologue;								
KW	phorbolin I homologue; cancer; tumour; autoimmune disorder;								
KW	circulatory system disorder; hypercholesterolaemia; viral infection;								
KW	neurological disease; neurofibromatosis; transcript editing; detection;								
ss.									
OS	Homo sapiens.								
XX									
Key	Location/Qualifiers								
FH	1..573								
CDS	/*tag= a								
FT	/product= "Human REE-2"								
FT									
XX									
PN	US6087108-A.								
XX									
PD	11-JUL-2000.								
XX									
PF	03-AUG-1998; 98US-0128395.								

RESULT 14

WPI; 2000-531340/48.
P-PSDB; AAB11973.

Detecting polynucleotide encoding human RNA editing enzyme comprising hybridizing an isolated and purified polynucleotide complementary to the polynucleotide and detecting the hybridization complex -
Claim 5; Fig 1A-B; 27pp; English.

This sequence represents the cDNA encoding human RNA editing enzyme REE-2. cDNA encoding REE-2 was initially isolated in a prostate tumour cDNA library, with the present sequence representing a consensus. REE-2 has chemical and structural homology with the human apob mRNA editing protein HEPR (28% identity), the rat HEPR homologue REPR (30% identity), and a portion of the mRNA editing enzyme phorbol I (43% identity). REE-2 was found to be expressed in a variety of cDNA libraries, a high proportion of which were derived from tumours, neuronal tissues, immune system cells or synovial tissue from arthritis patients. REE-2 is therefore thought to be associated with the development of cancer, autoimmune disorders, circulatory system disorders (e.g. hypercholesterolaemia), viral infections and neurological diseases (e.g., neurofibromatosis). REE-2 or its nucleic acids may be used in the diagnosis, treatment and prevention of such diseases via the modulation of transcript editing, which in turn has effects on the encoded protein (e.g., an alteration in protein activity). The invention specifically relates to methods of detecting nucleic acids encoding human REE-2 in a biological sample.

Sequence 610 BP: 158 A: 154 C: 154 G: 144 T: 0 other:

Query Match	19.2%;	Score 114.8;	DB 21;	Length 610;
Best Local Similarity	53.9%;	Prod. No. 5.9e-24;		
Matches	289;	Conservative	0;	Mismatches 232; Indels 15; Gaps 2;
Qy	16	ATGAACGGAGGAGTTTCTTTACCAATTCAAAAATGTCGGCTGAGGCTAAGGTCGGCGT	75	
Db	34	ATGATATCCAGGCACATTCTACTTCCAATTTAAAAACCTATGGNAGCCACGATCGGAAC	93	
Qy	76	GAGACCTTACCTTGCTACTGCTAGTGAAGAGCGGTGACAGTGTACATCTTTTCTACTGGAC	135	
Db	94	GAAACTTGGCTGTGCTTCCCGTGAAGGTATAAAGCGCGCTCAGTTGTCTCTCTGGAAG	155	
Qy	136	TTTGGTTATCTTCGCAA-----TAAGAACGGCTGCCACGTGCGAATGCTCTTCT	183	
Db	154	ACGGGCTCTTCGAAACACAGTGGATTTCTGAGACCCATTGTCTATGCAGAAAGGTGCTT	213	
Qy	184	CTCGCTTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCCGCTCACCTGGTTC	243	
Db	214	CTCTCTTGGTTCTCGAGACATACTGTCTCTTAACAACAAAGTACACAGGTACCTTGGTAC	273	
Qy	244	ACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTCTTCTGCGAGGAAC	303	
Db	274	ACATCTTGGAGCCCTTCCCGAGACTGTGACGGGAGGTGGCCGAGTTCTTGGCCAGGCAC	333	
Qy	304	CCCAACTCTAGTCTGAGGATCTTACC CGGGCCCTCTACTTCTGTGAGGACCCGCAAGCT	363	
Db	334	AGCAACGTGAATCTCAGCATCTTTCAGCGCCGCTCTACTACTTCCCAATTC---CATGT	390	
Qy	364	GAGCCCGAGGGCTGCGGGGCTCCACCGCGCGGGGTGCAAAATAGCATCATCACCTTC	423	
Db	391	TACCAGAGGGGCTCCCGAGCCTCAGTCAAGGAGGGTCCGCTGTGGACATCATGGACTAT	450	
Qy	424	AAAGATTATTTTACTGTCTGGAATACTTTTCTAGAAAAACCATGAAAGAACTTTCAAAGCC	483	
Db	451	GAAATTTTAAATATTGTTGGGAAACTTTGTGTACAATGATAATGACCCATTCGAAGCT	510	
Qy	484	TGGGAAGGGCTGCATGAAATTCAGTTGCTCTCTCCAGACAGCTTTCGGCGCATCCT	539	
Db	511	TGGGAAGGGATTAAAAACCAACTTTTCGACTCTGAAAAAGAGGCTACGGGAGAGTCT	566	

RESULT 14

AAS41420
ID AAS41420 standard; cDNA: 950 BP.
XX
AC AAS41420;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #636.
KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
OS
XX Homo sapiens.
XX
PN WO200155301-A2.
XX
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-465566/50.		
DR	P-PSDB; AAU23550.		
XX	Novel polypeptides and polynucleotides useful for diagnosing,		
PT	preventing, treating neural, immune system, muscular, reproductive,		
PT	pulmonary, cardiovascular, renal, proliferative disorders and cancerous		
PT	diseases -		
XX	Claim 4; SEQ ID No 646; 1180pp; English.		
XX	The present invention relates to the isolation of novel human enzyme		
CC	polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences		
CC	encoding them. The enzyme polypeptides of the invention may comprise the		
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,		
CC	isomerases or ligases. The sequences of the invention are useful in the		
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of		
CC	disorders including hyperproliferative disorders (e.g. cancer),		
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders		
CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),		
CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders		
CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),		
CC	blood-related disorders (e.g. haemophilia), reproductive disorders		
CC	(e.g. infertility) and infectious disorders (e.g. Influenza). The		
CC	polynucleotides of the invention can also be used in gene therapy.		
CC	AAU40785-AAU41684 represent cDNA sequences encoding for the novel human		
CC	enzyme polypeptides of the invention.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence 950 BP; 219 A; 289 C; 219 G; 220 T; 3 other;		
SQ	Query Match 19.2%; Score 114.8; DB 22; Length 950;		
	Best Local Similarity 53.9%; Pred. No. 7.3e-24;		
	Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;		
QY	16 ATGAACCGGAGGAGTTCTTTTACCAATTCAAAATGTCGCTGAGGCTAGGTCGCGT 75		
DB	95 ATGTATCCAGGCACATTCTACTTCCATTTTAAAAACCTATGGAAGCCACGATCGGAAC 154		
QY	76 GAGACCTACCTGTGCTAGTGTAGTGAAGCGGTGACAGTGTCTATCTTTTCTACCTGCAC 135		
DB	155 GAAACTTGGTGTGCTTCCACCGTGGAGAGGTATTAAGCGCCGCTCAGTTGCTCTCTGAAG 214		
QY	136 TTTGGTTATCTTCCAA-----TAAGAACGGCTGCCACGTGGAAATGCTCTTC 183		
DB	215 ACGGGCGTCTCCGAACAGGTGGATTCTGAGACCCATTGTCATGCAGAAAGGTGCTTC 274		
QY	184 CTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCTACCTGCTTC 243		
DB	275 CTCCTTTGTTCTGCGACGACATATCTCTCTTAACACAAAGTACCACTGCTGTAC 334		
QY	244 ACCTCCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGCGACTTTCTCGGAGGAAC 303		
Db	335 ACATCTTGGAGCCCTTCCCCAGACTGTGCAGGGAGGTGGCCGAGTTCTTGCCAGGCAC 394		
Qy	304 CCAACCTCAGTCTGAGGATCTTCCAGCGCGCTCTACTTCTGTGAGGACCGCAAGGCT 363		
Db	395 AGCAAGCTGAATCTCACCATTCTTACCGCCGCTCTTACTACTTCCAGTATC---CATGT 451		
Qy	364 GAGCCCGAGGGCTGGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTC 423		
Db	452 TACCAGGAGGGCTCCCGACGCTGAGTCAGAAAGGGTCTGCTGTGGAGATCATGGACTAT 511		
Qy	424 AAAGATTATTTTACTCTCTGGAATACCTTTTGTAGAAACCATGAAACATTTTCAAGCC 483		
Db	512 GAAGATTTTAAATATTGTTGGGAAACCTTTTGTGTACAAATGATATGAGCCATTCAAGCCT 571		
Qy	484 TGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCT 539		
Db	572 TGAAGGGATTAAAAACCACTTTTCGACTTCTGAAAAGAGGCTACGGGAGAGTCT 627		
RESULT 15			
AAF16264			
ID	AAF16264 standard; cDNA: 987 BP.		
XX	AC AAF16264;		
XX	13-MAR-2001 (first entry)		
DT	Human prostate cancer antigen nucleotide sequence SEQ ID NO:699.		
XX	Human: prostate cancer; prostate cancer antigen; detection; diagnosis;		
XX	neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;		
XX	vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;		
XX	antibacterial; gene therapy; neural; immune; reproductive; renal;		
XX	gastrointestinal; pulmonary; cardiovascular; proliferative disorder;		
XX	wound; infectious disease; ss.		
OS	Homo sapiens.		
XX	WO200055174-A1.		
PN	21-SEP-2000.		
PD	08-MAR-2000; 2000WO-US05988.		
XX	12-MAR-1999; 99US-0124270.		
PR	(HUMA-) HUMAN GENOME SCI INC.		
XX	(ROSE/) ROSEN C A.		
PA	Rosen CA, Ruben SM;		
XX	WPI; 2000-587513/55.		
DR	P-PSDB; AAB57061.		
XX	Prostate cancer associated gene sequences, referred to as prostate		
PT	cancer antigens, useful for treatment, prevention, and diagnosis of		
PT	disorders such as prostate cancer -		
XX	Claim 1; Page 1129; 2338pp; English.		
PS	AAF15566 to AAF16505 encode the human prostate cancer associated		
XX	proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.		
CC	The prostate cancer antigens can have neuroprotective, cytostatic,		
CC	cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,		
CC	nephrotropic, antiinfective, gynaecological and antibacterial activities,		
CC	and can be used in gene therapy. The prostate cancer antigen		
CC	polynucleotides may be used for detection of prostate cancer, chromosome		
CC	identification, as chromosome markers, and for numerous other diagnostic		
CC	or research purposes. The prostate cancer antigens may be used to treat		
CC	disorders such as neural, immune, muscular, reproductive,		
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative		
CC	disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to		
CC	AAB57303 represent sequences used in the exemplification of the present		

CC	invention.	
XX		
SQ	Sequence 987 BP; 231 A; 298 C; 232 G; 224 T; 2 other;	
	Query Match 19.2%; Score 114.8; DB 21; Length 987;	
	Best Local Similarity 53.9%; Pred. No. 7.4e-24;	
	Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;	
QY	16 ATGAACCGGAGGAACCTTCTTACCAANTTCRAAATATGCCGTGGCTAAGGGTCGGCGT 75	
DB	132 ATGTATCAGGCACATCTACTTCCAAATTTAAAAACCTATGGGAAGCCAAACGATCGGAAC 191	
QY	76 GAGACTACCTGTGTAGTGTGAAGAGGCGTGACAGTGTACATCCCTTTTCACTGGAC 135	
DB	192 GAAACTTGGCTGTGCTTCACCGTGAAGGTATAAAGCGCGCTCAGTTGTCTCCTGGAAG 251	
QY	136 TTTGGTTATCTCGCAA-----TAAGAAGGCTGCCAGCTGGAATTGCTCTTC 183	
DB	252 ACGGGCGTCTTCCGAACACGAGTGGATTCTGAGACCCATTGTCTATCGAAGAGTGCCTTC 311	
QY	184 CTCGCGTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGGTTC 243	
DB	312 CTCTCTTGGTCTGCGAGCAGATCTGTCTCTTAACAAAGTACCAGGTACCTGGGTAC 371	
QY	244 ACCTCCTGGAGCCCTGTGTACGACTGTGCCCGACATGTGGCGGACTTTCTGCGAGGGAAC 303	
DB	372 ACATCTTGAGGCCCTTGCCACACTGTGCAGGGGAGGTGGCCGAGTCTCTGCCCAGGCAC 431	
QY	304 CCCAACCTCAGTCTGAGGATCTTACCGGCGGCCCTCTACTTCTGTGAGGACCGCAAGGCT 363	
DB	432 AGCAACGTGAATCTCACCATCTTACCGCGCCGCTCTACTACTTCCAGTATC---CATGT 488	
QY	364 GAGCCCGAGGGGCTGGGGGCTGCACCGCGCGGGTGCAATAGCCCATCATGACCTTC 423	
DB	489 TACCAGGAGGGGCTCCGAGCCTGAGTCAGGAAGGGGTCGCTGTGGAGATCATGGACTAT 548	
QY	424 AAAGATTATTTTACTGTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAAAGCC 483	
DB	549 GAAGATTTAAATATTGTTGGGAAACCTTTGTGTACATGATATAGGCCATTCAGGCCT 608	
QY	484 TGGGAAGGCGTCGATGAAATTCAGTTCTCTCTCCAGACAGCTTCGGGCGCATCCT 539	
DB	609 TGGGAAGGGATTAAAAACCAACTTTCGACTTCTGAAAAAGAAAGGCTACGGGAGAGTCT 664	

Search completed: June 19, 2003, 00:37:11
Job time : 124.222 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:34:44 ; Search time 805.004 Seconds
(without alignments)
12010.756 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676
Perfect score: 597
Sequence: 1 atggacagcctcttgatgaa.....ttcgtactttgggactttga 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	743	12	BG686133 602638412
2	597	100.0	856	9	AL559877 602638412
3	597	100.0	872	12	BG758510 602712721
4	597	100.0	953	14	BQ065440 602712721
5	597	100.0	1052	14	BQ055935 602712721
6	581.8	97.5	820	12	BG757089 602715124

7	576.4	96.5	693	12	BG757392	BG757392	602711022
8	560.6	93.9	942	12	BF975166	BF975166	602244657
9	464	77.7	541	12	BF238155	BF238155	601811880
10	427.4	71.6	889	12	BG686876	BG686876	602850861
11	411.4	68.9	757	9	AJ446140	AJ446140	602850861
12	409.8	68.6	696	9	AJ453647	AJ453647	602850861
13	400	67.0	522	12	BG144705	BG144705	602850861
14	373	62.5	693	12	BF975096	BF975096	602245679
15	366.2	61.3	729	9	AJ450296	AJ450296	602245679
16	365.6	61.2	688	9	AJ450317	AJ450317	602245679
17	352	59.0	650	9	AJ449745	AJ449745	602245679
18	308.6	51.7	653	10	BB637360	BB637360	602245679
19	291.8	48.9	623	9	AJ450295	AJ450295	602245679
20	156.2	26.2	1034	12	BG755005	BG755005	602711511
21	146.8	24.6	843	12	BG758815	BG758815	602711511
22	118.4	19.8	713	10	BE409715	BE409715	601301573
23	118.4	19.8	751	10	BE383855	BE383855	601273177
24	118.4	19.8	833	12	BG281982	BG281982	602403168
25	117.8	19.7	737	12	BF686237	BF686237	602143689
26	117.8	19.7	884	13	BI834760	BI834760	603090364
27	116.4	19.5	1665	13	BM473129	BM473129	603090364
28	116.2	19.5	787	10	BE270707	BE270707	601185701
29	116	19.4	986	14	BQ877348	BQ877348	601185701
30	115.2	19.3	531	12	BF836372	BF836372	601185701
31	114.8	19.2	741	12	BE886229	BE886229	601509806
32	114.8	19.2	782	12	BE882870	BE882870	601509274
33	114.8	19.2	923	12	BG025943	BG025943	602292096
34	114.8	19.2	940	12	BG032405	BG032405	602301392
35	114.8	19.2	1005	13	BM474553	BM474553	602301392
36	114.8	19.2	1020	13	BM424095	BM424095	602301392
37	114.8	19.2	1063	14	BM809617	BM809617	602301392
38	114.8	19.2	1073	12	BG342101	BG342101	602463048
39	114.8	19.2	1100	14	BM914942	BM914942	602463048
40	114.8	19.2	1117	14	BM917194	BM917194	602463048
41	114.8	19.2	1119	11	BC021080	BC021080	602463048
42	114.8	19.2	1128	13	BM559604	BM559604	602463048
43	113.4	19.0	574	14	BM787460	BM787460	602463048
44	113.4	19.0	974	14	BQ707723	BQ707723	602463048
45	113.4	19.0	1041	14	BM922257	BM922257	602463048

ALIGNMENTS

RESULT 1
BG686133
LOCUS 602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5', linear EST 01-MAY-2001
DEFINITION mRNA sequence.
ACCESSION BG686133
VERSION BG686133.1 GI:13917530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 743)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1626 row: 9 column: 03
High quality sequence stop: 740.
Location/Qualifiers
1. .743
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

```

```

BASE COUNT      176 a   197 c   188 g   182 t
ORIGIN

```

```

Query Match      100.0%; Score 597; DB 12; Length 743;
Best Local Similarity 100.0%; Pred. No. 1.3e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
DB 78 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 137
QY 61 GCTAAGGTCGGCGTGAGACCTACCTGTGTACGTAGTGAAGAGGGGTGACAGTGTCTA 120
DB 138 GCTAAGGTCGGCGTGAGACCTACCTGTGTACGTAGTGAAGAGGGGTGACAGTGTCTA 197
QY 121 TCCTTTTCTACTGACCTTTGGTTATCTCGCAATTAAGAACGGCTGCCACGTGGAATTGCTC 180
DB 198 TCCTTTTCTACTGACCTTTGGTTATCTCGCAATTAAGAACGGCTGCCACGTGGAATTGCTC 257
QY 181 TTCTCTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTACCGCGTCACCTGG 240
DB 258 TTCTCTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTACCGCGTCACCTGG 317
QY 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 300
DB 318 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 377
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
DB 378 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 437
QY 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
DB 438 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 497
QY 421 TTCAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 480
DB 498 TTCAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 557
QY 481 GCCTGGGAGGGCTGATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCATCCCTT 540
DB 558 GCCTGGGAGGGCTGATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCATCCCTT 617
QY 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGAGCGCATTTCTGTTGGGACTTTGA 597
DB 618 TTGCCCTCTGATGAGGTTGATGACTTACGAGAGCGCATTTCTGTTGGGACTTTGA 674

```

```

RESULT 2
AL559877 LOCUS      AL559877 LTI_FL011_BCI Homo sapiens cDNA clone CS0DG003YB14 5 prime
DEFINITION      , mRNA sequence.
ACCESSION      AL559877
VERSION        AL559877.1 GI:12905793
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 856)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG003YB14"
/clone_lib="LTI_FL011_BCI"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned, into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center-Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

```

```

BASE COUNT      209 a   217 c   202 g   226 t   2 others
ORIGIN

```

```

Query Match      100.0%; Score 597; DB 9; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
DB 19 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 78
QY 61 GCTAAGGTCGGCGTGAGACCTACCTGTGTACGTAGTGAAGAGGGGTGACAGTGTCTA 120
DB 79 GCTAAGGTCGGCGTGAGACCTACCTGTGTACGTAGTGAAGAGGGGTGACAGTGTCTA 138
QY 121 TCCTTTTCTACTGACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAATTGCTC 180
DB 139 TCCTTTTCTACTGACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAATTGCTC 198
QY 181 TTCTCTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTCTACCGCGTCACCTGG 240
DB 199 TTCTCTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTCTACCGCGTCACCTGG 258
QY 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTTCTGCGAGGG 300
DB 259 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTTCTGCGAGGG 318
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
DB 319 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 378
QY 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
DB 379 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 438
QY 421 TTCAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 480
DB 439 TTCAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 498
QY 481 GCCTGGGAGGGCTGATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCATCCCTT 540
DB 499 GCCTGGGAGGGCTGATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGGCATCCCTT 558
QY 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGAGCGCATTTCTGTTGGGACTTTGA 597

```



```
Db 559 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 615

RESULT 3
BG758510 872 bp mRNA linear EST 15-MAY-2001
LOCUS 602712721F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853069 5',
DEFINITION mRNA sequence.
ACCESSION BG758510
VERSION BG758510.1 GI:14069163
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 872)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1698 row: i column: 06
High quality sequence stop: 836.
Location/Qualifiers
1. 872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4853069"
/clone_lib="NIH_MGC_48"
/tissue_type="Primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 211 a 221 c 212 g 228 t
ORIGIN
Query Match 100.0%; Score 597; DB 12; Length 872;
Best Local Similarity 100.0%; Pred. No. 1.4e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACACCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 60
Db 66 ATGGACACCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 125
QY 61 GCTAGGTCGCGGTGACACCTACTGCTGCTACGTAGTGAAGCGGTGACAGTCTACA 120
Db 126 GCTAAGGTCGCGGTGACACCTACTGCTGCTACGTAGTGAAGCGGTGACAGTCTACA 185
QY 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGACAGTGAATTCGTC 180
Db 186 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGACAGTGAATTCGTC 245
QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGGTGCTACCGGCTCACCTGG 240
Db 246 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGGTGCTACCGGCTCACCTGG 305
QY 241 TTCACCTCTGGAGCCCTCTGTACAGCTGTGCCGACATGTGGCCGACTTCTTCGGAGGG 300
Db 306 TTCACCTCTGGAGCCCTCTGTACAGCTGTGCCGACATGTGGCCGACTTCTTCGGAGGG 365

QY 301 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 366 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 425
QY 361 GCTGAGCCCGAGGGGTGCGGGGTGCACCGCGCGGGGTGCAGAAATAGCCATCATGACC 420
Db 426 GCTGAGCCCGAGGGGTGCGGGGTGCACCGCGCGGGGTGCAGAAATAGCCATCATGACC 485
QY 421 TTCAAGATATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTTCAA 480
Db 486 TTCAAGATATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTTCAA 545
QY 481 GCCTGGGAAGGCGTCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 546 GCCTGGGAAGGCGTCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTT 605
QY 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 597
Db 606 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 662

RESULT 4
BQ065440 953 bp mRNA linear EST 02-APR-2002
LOCUS AGENCOURT_6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
DEFINITION 5', mRNA sequence.
ACCESSION BQ065440
VERSION BQ065440.1 GI:19894486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 953)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2108 row: p column: 10
High quality sequence stop: 634.
Location/Qualifiers
1. 953
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5929977"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 238 a 236 c 233 g 246 t
ORIGIN
Query Match 100.0%; Score 597; DB 14; Length 953;
Best Local Similarity 100.0%; Pred. No. 1.5e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGGACAGCCTCTTGATGTAACCGGAGGAAGTTTCTTTTACCAATTCAAAAATGTCCTGG 60
Db 64 ATGGACAGCCTCTTGATGTAACCGGAGGAAGTTTCTTTTACCAATTCAAAAATGTCCTGG 123
QY 61 GCTAAGGTCGGCGTCAGACCTACCTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 120
Db 124 GCTAAGGTCGGCGTCAGACCTACCTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 183
QY 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAAGAGGCTGCCACCTGGAATTTGCTC 180
Db 184 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAAGAGGCTGCCACCTGGAATTTGCTC 243
QY 181 TTCCTCCGCTACATCTCGGACCTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACCTGG 240
Db 244 TTCCTCCGCTACATCTCGGACCTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACCTGG 303
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 300
Db 304 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 363
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 364 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
QY 361 GCTGAGCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAATAGCCATCATGACC 420
Db 424 GCTGAGCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAATAGCCATCATGACC 483
QY 421 TTCAAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAATTTTCAAA 480
Db 484 TTCAAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAATTTTCAAA 543
QY 481 GCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 544 GCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCAGACAGCTTCGGCGCATCCTT 603
QY 541 TTGCCCTCTATGAGTTGATGACTTACGAGCGCATTTTCGTAATTTGGGACTTTGA 597
Db 604 TTGCCCTCTATGAGTTGATGACTTACGAGCGCATTTTCGTAATTTGGGACTTTGA 660
```

```
RESULT 5
BQ055935
LOCUS
DEFINITION
AGENCOURT_6796291 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5808181
5', mRNA sequence.
ACCESSION
BQ055935
VERSION
BQ055935.1 GI:19815262
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1052)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2051 row: m column: 14
High quality sequence stop: 665.
FEATURES
Location/Qualifiers
1..1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5808181"
```

```
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pORF7; Site_1: xhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 276 a 250 c 258 g 266 t 2 others
ORIGIN
Query Match 100.0%; Score 597; DB 14; Length 1052;
Best Local Similarity 100.0%; Pred. No. 1.6e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGACAGCCTCTTGATGTAACCGGAGGAAGTTTCTTTTACCAATTCAAAAATGTCCTGG 60
Db 64 ATGGACAGCCTCTTGATGTAACCGGAGGAAGTTTCTTTTACCAATTCAAAAATGTCCTGG 123
QY 61 GCTAAGGTCGGCGTCAGACCTACCTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 120
Db 124 GCTAAGGTCGGCGTCAGACCTACCTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 183
QY 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAAGAGGCTGCCACCTGGAATTTGCTC 180
Db 184 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAAGAGGCTGCCACCTGGAATTTGCTC 243
QY 181 TTCCTCCGCTACATCTCGGACCTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240
Db 244 TTCCTCCGCTACATCTCGGACCTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 303
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 300
Db 304 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 363
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 364 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
QY 361 GCTGAGCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAATAGCCATCATGACC 420
Db 424 GCTGAGCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAATAGCCATCATGACC 483
QY 421 TTCAAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAATTTTCAAA 480
Db 484 TTCAAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAATTTTCAAA 543
QY 481 GCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 544 GCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCAGACAGCTTCGGCGCATCCTT 603
QY 541 TTGCCCTCTATGAGTTGATGACTTACGAGCGCATTTTCGTAATTTGGGACTTTGA 597
Db 604 TTGCCCTCTATGAGTTGATGACTTACGAGCGCATTTTCGTAATTTGGGACTTTGA 660
RESULT 6
BQ757089
LOCUS
DEFINITION
602715124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
mRNA sequence.
ACCESSION
BQ757089
VERSION
BQ757089.1 GI:14067742
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 820)
```

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI704 row: o column: 06
High quality sequence stop: 675.
Location/Qualifiers
1. 820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:485517"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
210 a 230 c 200 g 180 t

FEATURES
source

BASE COUNT 210 a 230 c 200 g 180 t
ORIGIN
Query Match 97.5%; Score 581.8; DB 12; Length 820;
Best Local Similarity 99.5%; Pred. No. 5.2e-164;
Matches 594; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 ATGGACAGCCTCTTCATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 60
DB 85 ATGGACAGCCTCTTCATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 144
QY 61 GCTAAGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
DB 145 GCTAAGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 204
QY 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCTC 180
DB 205 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCTC 264
QY 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGACCCCTGGCGCTGCTACCGGTACCTGG 240
DB 265 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGACCCCTGGCGCTGCTACCGGTACCTGG 324
QY 241 TTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTCTGCGAGGG 300
DB 325 TTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTCTGCGAGGG 384
QY 301 AACCCCAACCTCAGTCTGAGNATCTTACCGCGCGCTCTACTTCTGTGAGAGCCGCAAG 360
DB 385 AACCCCAACCTCAGTCTGAGNATCTTACCGCGCGCTCTACTTCTGTGAGAGCCGCAAG 444
QY 361 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420
DB 445 GCTGAGCCGAGGGCTCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 504
QY 421 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 480
DB 505 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 564
QY 481 GCCTGGGAGGGCTGCATGAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 540
DB 565 GCCTGGGAGGGCTGCATGAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 624

QY 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTGCTACTTTGGACTTTGCA 597
DB 625 TTGCCCTGTATGAGTTGATGACTTACGAGACGCA-TTCGTACTTAGGACTTCCA 680
RESULT 7
BG757392
LOCUS 602711022F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851580 5',
DEFINITION mRNA sequence.
ACCESSION BG757392
VERSION BG757392.1 GI:14068045
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI694 row: k column: 05
High quality sequence stop: 693.
Location/Qualifiers
1. 693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4851580"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
157 a 188 c 178 g 170 t

FEATURES
source

BASE COUNT 157 a 188 c 178 g 170 t
ORIGIN
Query Match 96.5%; Score 576.4; DB 12; Length 693;
Best Local Similarity 99.7%; Pred. No. 2e-162;
Matches 588; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATGGACAGCCTCTTCATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 60
DB 81 ATGGACAGCCTCTTCATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 140
QY 61 GCTAAGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 119
DB 141 GCTAAGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 200
QY 120 ATCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCT 179
DB 201 ATCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCT 260
QY 180 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCCCTGCTACCGGCTCACCTG 239
DB 261 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCCCTGCTACCGGCTCACCTG 320

```

QY      240 GTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 299
      |||
Db      321 GTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 380
QY      300 GAACCCCAACCTAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACGCAA 359
      |||
Db      381 GAACCCCAACCTAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACGCAA 440
QY      360 GGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCAATAGCCATCATGAC 419
      |||
Db      441 GGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCAATAGCCATCATGAC 500
QY      420 CTTCAAGATTATTTTACTGCTGAATACTTTGTAGAAAACCATGAAGAACTTCA 479
      |||
Db      501 CTTCAAGATTATTTTACTGCTGAATACTTTGTAGAAAACCATGAAGAACTTCA 560
QY      480 AGCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGCGCATCCT 539
      |||
Db      561 AGCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGCGCATCCT 620
QY      540 TTTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGACTTTGG 589
      |||
Db      621 TTTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGACTTTGG 670

RESULT 8
LOCUS   BF975166
DEFINITION 602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4335639 5',
      mRNA sequence.
ACCESSION BF975166
VERSION   BF975166.1 GI:12342381
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 942)
          NIH-MGC http://mhc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapps-femail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LNCM1207 row: a column: 16
          High quality sequence stop: 707.
FEATURES             Location/Qualifiers
     1..942
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:4335639"
     /clone_lib="NIH_MGC_48"
     /tissue_type="primary B-cells from tonsils (cell line)"
     /lab_host="DH10B (phage-resistant)"
     /notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
     Site:2: EcoRI; cDNA made by oligo-dT priming.
     Directionally cloned into EcoRI/XhoI sites using the
     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
     for average insert size 1.8kb. Library constructed by Ling
     Hong in the laboratory of Gerald M. Rubin (University of
     California, Berkeley) using ZAP-cDNA synthesis kit
     (Stratagene) and Superscript II RT (Life Technologies).
     Note: this is a NIH_MGC Library."
     249 a 209 c 251 g 233 t

BASE COUNT      249 a 209 c 251 g 233 t
ORIGIN

Query Match      93.9%; Score 560.6; DB 12; Length 942;

```

```

Best Local Similarity 99.0%; Pred. No. 1.4e-157;
Matches 585; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY      1 ATGGACAGCCCTTTCATGAACCGGAGGAGTTTCTTTTACCAA-TTCAAAATATGTCGCTG 59
      |||
Db      78 ATGGACAGCCCTTTCATGAACCGGAGGAGTTTCTTTTACCAAATATGTCGCTG 137
QY      60 GCCTAAGGCTCGGCGTGAGACTACCTGCTACGTAGTGAAGAGGCTGACAGTGCTAC 119
      |||
Db      138 GCCTAAGGCTCGGCGTGAGACTACCTGCTAGTGAAGAGGCTGACAGTGCTAC 197
QY      120 ATCTTTTCACTGAGACTTTGTTATCTTCGCAATAAGAACGCTGCCACGTGGAAATGCT 179
      |||
Db      198 ATCTTTTCACTGAGACTTTGTTATCTTCGCAATAAGAACGCTGCCACGTGGAAATGCT 257
QY      180 CTTCTCCGCTACATCTCGGACTCGGACTAGACCTTGGCCGCTGCTACCGCGTCACCTG 239
      |||
Db      258 CTTCTCCGCTACATCTCGGACTCGGACTAGACCTTGGCCGCTGCTACCGCGTCACCTG 317
QY      240 GTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCCGACATGTGCCGACTTCTTCGCGAGG 299
      |||
Db      318 GTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGACTTCTTCGCGAGG 377
QY      300 GAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAA 359
      |||
Db      378 GAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAA 437
QY      360 GGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCAATAGCCATCATGAC 419
      |||
Db      438 GGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCAATAGCCATCATGAC 497
QY      420 CTTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAACTTTC 479
      |||
Db      498 CTTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAACTTTC 557
QY      480 AGCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGCGCATCCT 539
      |||
Db      558 AGCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGCGCATCCT 617
QY      540 TTTGCCCTGTATGA-GGTTGATGACTTACGAGACGCAATTTCTGACTTTGG 589
      |||
Db      618 TTGGGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTGG 668

RESULT 9
LOCUS   BF238155
DEFINITION 601811880F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054915 5',
      mRNA sequence.
ACCESSION BF238155
VERSION   BF238155.1 GI:11152074
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 541)
          NIH-MGC http://mhc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapps-femail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LNCM895 row: p column: 20
          High quality sequence stop: 541.
FEATURES             Location/Qualifiers
     1..541

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4054915"
/clone_lib="NIH_MGC_48"
/tissue_type="Primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/site="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 120 a 150 c 142 g 129 t
ORIGIN
Query Match 77.7%; Score 464; DB 12; Length 541;
Best Local Similarity 100.0%; Pred. No. 1.3e-128;
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCTGG 60
Db |
QY 78 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCTGG 137
Db |
QY 61 GCTAAGGGTCGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db |
QY 138 GCTAAGGGTCGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 197
Db |
QY 121 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTCCTC 180
Db |
QY 198 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTCCTC 257
Db |
QY 181 TTCTCCCGTACATCTCGACCTGGGACCTAGACCCCTGGCCGCTGCTACCGCTCAGCTGG 240
Db |
QY 258 TTCTCCCGTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCTCAGCTGG 317
Db |
QY 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTCTTGGGAGG 300
Db |
QY 318 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTCTTGGGAGG 377
Db |
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db |
QY 378 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 437
Db |
QY 361 GCTGAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAGATAGCCATCATGACC 420
Db |
QY 438 GCTGAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAGATAGCCATCATGACC 497
Db |
QY 421 TTCAAGATATTTTACTGCTGGATATCTTTGTAGAAAACCA 464
Db |
QY 498 TTCAAGATATTTTACTGCTGGATATCTTTGTAGAAAACCA 541

```

```

RESULT 10
BG686876 889 bp mRNA linear EST 01-MAY-2001
LOCUS 602650861F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4763247 5',
DEFINITION mRNA sequence.
ACCESSION BG686876
VERSION BG686876.1 GI:13918273
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

```

```

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1618 row: j column: 16
High quality sequence start: 6
High quality sequence stop: 727.
FEATURES
Location/Qualifiers
1..889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4763247"
/clone_lib="NIH_MGC_48"
/tissue_type="Primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/site="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 205 a 223 c 257 g 203 t 1 others
ORIGIN
Query Match 71.6%; Score 427.4; DB 12; Length 889;
Best Local Similarity 99.8%; Pred. No. 1.8e-117;
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCTGG 60
Db |
QY 38 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCTGG 97
Db |
QY 61 GCTAAGGGTCGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db |
QY 98 GCTAAGGGTCGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 157
Db |
QY 121 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTCCTC 180
Db |
QY 158 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTCCTC 217
Db |
QY 181 TTCTCCCGTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCTCAGCTGG 240
Db |
QY 218 TTCTCCCGTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCTCAGCTGG 277
Db |
QY 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTCTTCTGCGAGGG 300
Db |
QY 278 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTCTTCTGCGAGGG 337
Db |
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db |
QY 338 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 397
Db |
QY 361 GCTGAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAGATAGCCATCATGACC 420
Db |
QY 398 GCTGAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAGATAGCCATCATGACC 457
Db |
QY 421 TTCAAGAT 429
Db |
QY 458 TTCAAGAT 466

```

```

RESULT 11
AJ446140
LOCUS AJ446140 riken1 Gallus gallus cDNA clone 12j9rl, mRNA sequence.
DEFINITION AJ446140
ACCESSION AJ446140
VERSION AJ446140.1 GI:20213361

```

```
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 757)
AUTHORS Buerstedde J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
COMMENT Contact: Buerstedde JM
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
    source
        1..757
            Location/Qualifiers
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /clone="12j9rl"
                /clone_lib="rikenl"
                /cell_type="bursal lymphocyte"
                /dev_stage="2-3 weeks old"
                /note="CB inbred strain"
BASE COUNT 180 a 205 c 184 g 188 t
ORIGIN
Query Match 68.9%; Score 411.4; DB 9; Length 757;
Best Local Similarity 80.6%; Pred. No. 1.1e-112;
Matches 481; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
Qy 1 ATGCACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
Db 97 ATGCACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 156
Qy 61 GCTAAGGGTGGCGTGGAGACCTACCTGTGTACGTAGTGAAGAGCGCTGACAGTGCTACA 120
Db 157 GCCAAGGGCGTGGTGAACCTACCTCTGTATGTTGTGAAGCGCGTGAAGTGCTACA 216
Qy 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 180
Db 217 TCATGCTCCCTGGACTTTGGATACCTGCGTAACAAGATGGTTGCCATGTGGAGTTCTC 276
Qy 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGCTGCTACCGCGTCACCTGG 240
Db 277 TTCCTAGCTACATCTCAGCTTGGGACCTGGACCGGCGCGCTGCTACCGCATCACATGG 336
Qy 241 TTCACCTCCGAGCCCGCTCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 300
Db 337 TTCACCTCCGAGCCCGCTCTATGACTGTGCCGACATGTGGCTGACTTCTCTGTCGCC 396
Qy 301 AACCCCAACCTCAGTCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 397 TACCCCAACTTGACCTTCGCACTTTTCACTGCCCGCTCTACTTCTGTGAAGATCGCAAG 456
Qy 361 GCTGAGCCGAGGGGTGGCGGCTGACCGCGCGGGTGCAGAAATAGCCATCATGACC 420
Db 457 GCTGAGCGTGAAGGGCTGAGACGCCCTGACCGCGCGTGGGCCCAATGCGCATGACT 516
Qy 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
Db 517 TTCAAAGATTCTTCTACTGCTGGACACCTTTGTGGAGAACAGGAAAGACATTCAAA 576
Qy 481 GCTGGGAAGGGCTGATGAAATTCAGTTTCGTCCTCCAGACAGCTTCGGGCGATCCCTT 540
Db 577 GCTGGGAAGGGCTGATGAAATTCAGTTTCGTCCTCCAGACAGCTTCGGGCGATCCCTT 636
Qy 541 TTGCCCTGTATGAGTTGACTGACTTACGAGACGCAATTTGCTGTTGGACTTTGA 597
Db 637 CTGCCACTGTATGAAGTAGATGATTACGAGATGCTTTAAACTCTGGGACTTTGA 693
RESULT 12
```

```
AJ453647
LOCUS chicken.
DEFINITION AJ453647 rikenl Gallus gallus cDNA clone 35b23r1, mRNA sequence.
ACCESSION AJ453647
VERSION AJ453647.1 GI:20263743
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 696)
AUTHORS Buerstedde J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
    source
        1..696
            Location/Qualifiers
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /clone="35b23r1"
                /clone_lib="rikenl"
                /cell_type="bursal lymphocyte"
                /dev_stage="2-3 weeks old"
                /note="CB inbred strain"
BASE COUNT 160 a 193 c 173 g 170 t
ORIGIN
```

```
Query Match 68.6%; Score 409.8; DB 9; Length 696;
Best Local Similarity 80.4%; Pred. No. 3.2e-112;
Matches 480; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
Qy 1 ATGCACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
Db 97 ATGCACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 156
Qy 61 GCTAAGGGTGGCGTGGAGACCTACCTGTGTACGTAGTGAAGAGGGGTGACAGTGCTACA 120
Db 157 GCCAAGGGCGTGGTGAACCTACCTCTGTATGTTGTGAAGCGCGTGAAGTGCTACA 216
Qy 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 180
Db 217 TCATGCTCCCTGGACTTTGGATACCTGCGTAACAAGATGGTTGCCATGTGGAGTTCTC 276
Qy 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGCTGCTACCGCGTCACCTGG 240
Db 277 TTCCTAGCTACATCTCAGCTTGGGACCTGGACCGCGCGCTGCTACCGCATCACATGG 336
Qy 241 TTCACCTCTCGAGCCCGCTCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 300
Db 337 TTCACCTCTCGAGCCCGCTGTATGACTGTGCCGACATGTGGCTGACTTCTCTGTCGCC 396
Qy 301 AACCCCAACCTCAGTCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 397 TACCCCAACTTGACCTTCGCGCATTTTCACTGCCCGCTCTACTTCTGTGAAGTTCGCAAG 456
Qy 361 GCTGAGCCGAGGGGTGGCGGCTGACCGCGCGGGGTGCAGAAATAGCCATCATGACC 420
Db 457 GCTGAGCGTGAAGGGCTGAGACGCCCTGACCGCGCGTGGGCCCAATGCGCATGACT 516
Qy 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
Db 517 TTCAAAGATTCTTCTACTGCTGGACACCTTTGTGGAGAACAGGAAAGACATTCAAA 576
Qy 481 GCTGGGAAGGGCTGATGAAATTCAGTTTCGTCCTCCAGACAGCTTCGGGCGATCCCTT 540
Db 577 GCTGGGAAGGGCTGATGAAATTCAGTTTCGTCCTCCAGACAGCTTCGGGCGATCCCTT 636
Qy 541 TTGCCCTGTATGAGTTGACTGACTTACGAGACGCAATTTGCTGTTGGACTTTGA 597
```

QY	394	GC	GGGGTGC	CAAAATAG	CCATCAT	CAGCACTT	CAAGAC	TATATTTT	TACTGCTG	GAATAC	TTTT	451						
Db	311	GCT	GGGTC	CAGAT	TCGGGAT	CAT	CACCTT	CAAGAC	TATATTTT	TACTGCTG	GAATAC	370						
QY	454	GT	AAAAAC	CAATGA	AAAGACTT	TTC	AAAGCCT	GGGAAG	CGCTGC	ATGAA	AAATTC	AGTT	513					
Db	371	GT	AAAAAT	CGTGA	AAAGACTT	TTC	AAAGCCT	GGGAAG	CGCTGC	ATGAA	AAATTC	AGTT	430					
QY	514	CT	CTCCAG	ACAGCT	TCGCGGC	ATCCT	TTT	TGCCCC	TGTAT	GAGT	TTGAT	AGTT	573					
Db	431	CT	AAACAG	ACAACT	TCGCGGC	ATCCT	TTT	TGCCCC	TGTAT	GAGT	TTGAT	AGTT	490					
QY	574	GC	ATTTCG	TACTTT	GGGAC	TTTGA	597											
Db	491	GC	ATTTCG	TATG	TGTTGG	ATTGTA	514											
RESULT 14																		
BF975096																		
LOCUS																		
DEFINITION	602245679f1	NIH_MGC_48	Homo sapiens	CDNA clone	IMAGE:4336722	5'												
ACCESSION	BF975096																	
VERSION	BF975096.1	GI:12342311																
KEYWORDS	EST.																	
SOURCE	human.																	
ORGANISM	Homo sapiens																	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																	
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.																	
TITLE	1 (bases 1 to 693)																	
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .																	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cqabps-remail.nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LNCMI209 row: n column: 19 High quality sequence stop: 692. Location/Qualifiers																	
FEATURES																		
source	1..693																	
	/organism="Homo sapiens"																	
	/db_xref="taxon:9606"																	
	/clone="IMAGE:4336722"																	
	/tissue_type="NIH_MGC_48"																	
	/lab_host="DH10B (phage-resistant)"																	
	/note="Organ: B-cells; Vector: pOPB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library."																	
BASE COUNT	179 a	167 c	162 g	185 t														
ORIGIN																		
Query Match																		
Best Local Similarity	62.5%																	
Matches	373; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;										
QY	225	CT	ACCGG	CTC	ACCTG	TTACCT	CTCT	CGAG	CCCT	CTGCT	ACGAC	TGTC	CGC	ACAT	GTGGC	284		
Db	1	CT	ACCGG	CTC	ACCTG	TTACCT	CTCT	CGAG	CCCT	CTGCT	ACGAC	TGTC	CGC	ACAT	GTGGC	60		
QY	285	CG	ACTTT	TCT	CGAG	GGAA	CAAC	CCCA	ACCT	TCAGT	CTCT	GAGG	ATCT	TTC	ACCG	CGCCT	CTACT	344

```
Db 61 CGACTTTCTGCGAGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCCCTCTACTT 120
Qy 345 CTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGGCGGTGCACCGCGCGGGGTGCA 404
Db 121 CTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGGCGGTGCACCGCGCGGGGTGCA 180
Qy 405 AATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATACTTTTGTAGAAAACCA 464
Db 181 AATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATACTTTTGTAGAAAACCA 240
Qy 465 TGAAGAACCTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA 524
Db 241 TGAAGAACCTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA 300
Qy 525 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTGTGATGACTTACGAGACGCATTTGCTAC 584
Db 301 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTGTGATGACTTACGAGACGCATTTGCTAC 360
Qy 585 TTTGGGACTTTGA 597
Db 361 TTTGGGACTTTGA 373
```

RESULT 15

```
AJ450296
LOCUS 729 bp mRNA linear EST 19-APR-2002
DEFINITION riken1 Gallus gallus cdna clone 24m22r1, mRNA sequence.
ACCESSION AJ450296
```

```
VERSION AJ450296.1 GI:20217517
```

```
KEYWORDS EST.
```

```
SOURCE chicken.
```

ORGANISM

```
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
```

```
REFERENCE 1 (bases 1 to 729)
```

```
AUTHORS Buerstedde,J.M.
```

```
TITLE Gallus gallus bursal lymphocyte EST
```

```
JOURNAL Unpublished (2002)
```

```
COMMENT Contact: Buerstedde JM
```

```
Cellular Immunology
```

```
Heinrich-Pette-Institute
```

```
Martinistr. 52, 20251 Hamburg, Germany
```

```
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
```

```
Location/Qualifiers
```

FEATURES

```
source
1..729
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="24m22r1"
/clone_lib="riken1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
```

BASE COUNT 173 a 199 c 187 g 170 t

ORIGIN

```
Query Match 61.3%; Score 366.2; DB 9; Length 729;
Best Local Similarity 79.8%; Pred. No. 4.5e-99;
Matches 457; Conservative 0; Mismatches 113; Indels 3; Gaps 2;

Qy 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCGTGG 60
Db 157 ATGGACAGCCTCTTGATGAAGGAGGAGTTCCTTCCTACAAATTCAGAAGACCTCGCGCTGG 216
Qy 61 GCTAAGGTCGGCTGAGACTACCTGCTGCTAGTGAAGAGCGCTGACAGTGCTAC 120
Db 217 GCCAAGCGCTGCTGAACCTACCTCTGTATGTTGTAAGCGCCGCTGACAGTGCTACA 276
Qy 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGCTGCCACAGTGGAAATGCTC 180
Db 277 TCATGCTCCCTGGACTTTGGTACTCTGCGTAAACAAGATGGTTGCCATGTGGAGGTCTC 336
```

```
Qy 181 TTCTCCGCTACATCTCTCGGACTTGGGACCTAGACCCCTGGCCGCTGCTACCGGTCACCTGG 240
Db 337 TTCTACGCTACATCTCTAGCTTTGGGACTTGGGACCTAGGACCCAGCGCTGCTACCGCATCACATGG 396
Qy 241 TTCAACCTCCCTGGAGCCCTCTGCTAGACTGTGCCGACATGTGGCCGACTTTCTTCGAGGG 300
Db 397 TTCAACCTCCCTGGAGCCCTCTGTTATGACTGTGCCGACATGTGGCTGACTTCTCTTCGTGCC 456
Qy 301 AACCCCAACCTCAGCTTCTGAGGATCTTCAACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 457 TACCCAAACTTGACCTCCGCATTTTCACTGCCCGCTCTACTTCTGTGAAGATCGCAAG 516
Qy 361 GCTGAGCCGAGGGCTGCGGCGCTGCACGCGCGGGGTGCAAAATAGCCATCATGACC 420
Db 517 GCTGAGCTGAGGGGCTGAGACGCTGCACGCGCTGGGGCCCAATCGCCATCATGACT 576
Qy 421 TTCAAGATTAATTTTACTGCTGGAATACTTTTCTAGAAAACCATGAAAGAACCTTTCAA 480
Db 577 TTCAAGATTTCTTCTACTGCTGGAACACGTTTGTGGAGAACAGGGAAAGACATTTCAA 636
Qy 481 GCCTGGGAAGGGCTGTCATGAAAATTCAGTTTCGTCT--CTCCAGACAGCTTCGGCGCATCC 538
Db 637 GCCTGGGAAGGGCTGTCATGAAAACCTCTGTCCATCTTGTCCAGGAAACTCCCGACGATCC 696
Qy 539 TT-TTGGCCCTGTATGAGGTGTGATGACTTTACGA 570
Db 697 TTCTTGGCACTGTATGAAGTAGATGATTTACGA 729
```

Search completed: June 19, 2003, 11:07:28

Job time : 808.004 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:01:09 ; Search time 25.0728 seconds
(without alignments)
7302.178 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676

Perfect score: 597

Sequence: 1 atggacagcctcttgatga.....ttcgtactttgggactttga 597

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCtUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.8	19.2	610	US-08-816-241-2	Sequence 2, Appli
2	114.8	19.2	610	US-09-128-395-2	Sequence 2, Appli
3	63.6	10.7	891	US-08-687-895-2	Sequence 2, Appli
4	63.6	10.7	891	US-09-040-482-2	Sequence 2, Appli
5	60	10.1	879	US-08-158-682A-3	Sequence 3, Appli
6	56.4	9.4	879	US-08-158-682A-1	Sequence 1, Appli
7	56.4	9.4	879	US-08-015-203-1	Sequence 1, Appli
8	38	6.4	1929	US-09-380-420C-1	Sequence 1, Appli
9	37.8	6.3	922	US-09-189-060B-73	Sequence 1, Appli
10	36.6	6.1	4689	US-09-105-537-34	Sequence 34, Appli
11	36.6	6.1	36778	US-09-105-537-5	Sequence 5, Appli
12	36.6	6.1	38506	US-09-320-878-19	Sequence 19, Appli
13	36.4	6.1	7218	US-08-232-463-14	Sequence 14, Appli
14	35.4	5.9	2190	US-09-015-188-1	Sequence 1, Appli
15	35	5.9	80246	US-09-078-294-4	Sequence 4, Appli
16	35	5.9	80595	US-09-078-294-3	Sequence 3, Appli
17	34.6	5.8	1597	US-08-724-974A-1	Sequence 1, Appli
18	34.2	5.7	5159	US-08-146-930-3	Sequence 3, Appli
19	34.2	5.7	5159	US-08-458-240-3	Sequence 3, Appli
20	34.2	5.7	5159	PCt-US93-03993-3	Sequence 3, Appli
21	33.8	5.7	15872	US-09-105-537-1	Sequence 1, Appli
22	33.8	5.7	4403765	US-09-103-840A-2	Sequence 2, Appli
23	33.8	5.7	4411529	US-09-103-840A-1	Sequence 1, Appli
24	33.2	5.6	35081	US-08-752-760A-1	Sequence 1, Appli
25	33	5.5	4911	US-09-718-692-1	Sequence 1, Appli
26	33	5.5	4911	US-09-718-852-1	Sequence 1, Appli
27	33	5.5	4911	US-09-718-815-1	Sequence 1, Appli

28	32.8	5.5	36519	3	US-08-923-137-2	Sequence 2, Appli
29	32.6	5.5	1356	1	US-08-317-880-4	Sequence 4, Appli
30	32.6	5.5	1356	2	US-08-782-396-4	Sequence 4, Appli
31	32.6	5.5	3954	1	US-08-317-880-1	Sequence 1, Appli
32	32.6	5.5	3954	2	US-08-782-396-1	Sequence 1, Appli
33	32.2	5.4	796	4	US-09-280-116-107	Sequence 107, App
34	32.2	5.4	1943	1	US-08-282-305-7	Sequence 7, Appli
35	32	5.4	8878	1	US-08-759-444-2	Sequence 2, Appli
36	32	5.4	9880	3	US-08-680-897-1	Sequence 1, Appli
37	31.8	5.3	657	4	US-09-527-345-3	Sequence 3, Appli
38	31.4	5.3	1008	4	US-09-500-569-1	Sequence 1, Appli
39	31.4	5.3	1621	3	US-09-013-881-14	Sequence 14, Appli
40	31.4	5.3	4141	4	US-09-245-281-42	Sequence 42, Appli
41	31.4	5.3	4141	4	US-09-207-359B-42	Sequence 42, Appli
42	31.4	5.3	13842	4	US-09-105-537-30	Sequence 30, Appli
43	31.4	5.3	29629	4	US-09-729-995-3	Sequence 3, Appli
44	31.4	5.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
45	31.2	5.2	1392	3	US-08-957-063-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-816-241-2
; Sequence 2, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646833
US-08-816-241-2

Query Match 19.2%; Score 114.8; DB 1; Length 610;
Best Local Similarity 53.9%; Pred. No. 1.9e-25;
Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

[illegible]

154	CGCGGCTCTTCCGAAAC	AGGTGGATCTTGAGAC	CCATTGTCATGACGAAAG	GTGCTTC	213
184	CTCGCTACATCTCGGACT	GGGACCTAGACCC	TGTACGCGCTGTCACCTGGTTC		243
214	CTCTTTGGTTCTCGGACG	ACATCTCTTAACACA	AGGTACGAGTACCTGGTAC		273
244	ACCTCTGGAGCCCTGCT	ACGACTGTGCCCGACA	TGTGCCGACTTTCTCGAGGGGAC		303

214	CTCTCTGGTTCTCGACGACATCATGTCTCTTAACACAAGTACCAGGTCACTGGTAT	Db
244	ACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGCAC	QY
274	ACATCTTGGAGCCCTTCCCAAGCTGTGCAGGGAGGTGCCGAGTTCTTGCCACGCAC	Db
304	CCCAACCTCACTCTGAGGATCTTCAACGCGCCCTCACTCTCTGAGCAACCGAAGCT	QY

274	ACATCTTGGAGCCCTTCCCAACATGTCAGGAGGTGCGCGAGTCTCTGCCAGCAC	333
Db		
304	CCCAACTCAGTCGTAGAGATCTTACCGCGCGCTCTACTCTGTGAGACGCCAAGGT	363
QY		
334	AGCAAGTGAATCTACCACTCTTCCGCGCGCTCTCTACTTCCAGTATC---CATGT	390
Db		
364	GAGCCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACCTTC	423
QY		

334	AGCAACGGAATCTACACACTTCCAGCCGCGCTCTACTACTTCCAGTATC---CATGT	390
Db		
364	GAGCCGAGGGGTGCGCGCTCACGGCGGGGTCAANTAGCGATCATGACCTTC	423
QY		
391	TACCAGAGGGGCTCCGACGCTGACTCAGGAAGGGTCCCTGGGATCATGGACTAT	450
Db		
424	AAAGATTATTTTACTCTGAAATACITTTGTGAAAACCACTGAAGAAGCTTTCAAGGC	483
QY		

391	TACCAGGAGGCTCCGACGCTGGTCAGGAAGGGTCCTGTGGAGATCATGGACAT	450
Db		
424	AAGAGTATTTTACTGCTGGAATACITTTGTAAANAACCATGAAGAAGCTTTCAAGCC	483
Qy		
451	GAGATTTTAAATATTTGTGGAAACCTTTGTGTACAATGATATGAGGCCATTCAGGCCT	510
Db		
484	TGGGAAGGGCTGCATGAATAATCAGTTCTCTCCAGACAGGTTTCGGCGCATCTCT	539
Qy		

[illegible]

Db 511 TCGAAGGGNTTAARACCAACTTTCGACTTCTGAAAAGAAGGCTACGGGAGAGCTCT 566

RESULT 2
US-09-128-395-2
: Sequence 2, Application US/09128395
: Patent No. 6087108
: GENERAL INFORMATION:

```

US-09-128-395-2
: Sequence 2 Application US/09128395
: Patent No. 6087108
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Goli, Surya K.
: TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
: NUMBER OF SEQUENCES: 5

```

APPLICANT: Bandman, Olga
 APPLICANT: Goli, Surva K.
 TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:

```

```

STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0

```

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/128,395
FILING DATE:
CLASSIFICATION:

```

* CURRENT APPLICATION DATA:
 * APPLICATION NUMBER: US/09/128,395
 * FILING DATE:
 * CLASSIFICATION:
 * PRIOR APPLICATION DATA:
 * APPLICATION NUMBER: 08/816,241
 * FILING DATE:
 * ATTORNEY/AGENT INFORMATION:

PRIOR APPLICATION DATA: 08/816,241
 APPLICATION NUMBER: 08/816,241
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0239 US
 TELECOMMUNICATION INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0239 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:

Db	606	GGTGCAGCTCTTTCATGTGGGAGGAGCGGAGATCCAG---	GCTGCTCTGAAGAAGCTGAAG	662
QY	391	CGGCGCGGGTGCAAATAGCCATCATGACCTTCAAGATATATTTT	TACTGCTCGAATACT	450
Db	663	GAGGCTTGGCTGTAACTGGGCATCATGAAGCCCGAGGACTTCGA	ATATGTCTGGCAGAAT	722
QY	451	TTTGTAGAAACCATCAAGAAGCTTTTCAAGCGCTGGAGGGCTGC	ATGAAAT	504
Db	723	TTTGTGGAGCAAGAGGGTGAATCCAAGGCGCTTCACGCCCTGG	GAGGACAT	776

RESULT 5
US-08-158-682A-3
: Sequence 3, Application US/08158682A

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/015,203
FILING DATE: 19930209
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coolley, Ronald B.
REGISTRATION NUMBER: 27,187
REFERENCE/DOCKET NUMBER: ARCD:069
TELEPHONE: (312) 744-0090
TELEFAX: (312) 245-4961
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 879 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-015-203-1

Query Match 9.4%; Score 56.4; DB 1; Length 879;
Best Local Similarity 53.4%; Pred. No. 1.3e-07;
Matches 142; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 197 CGGACTGGACCTAGACCTGCGCGCTGCTACCGGTACCTGGTTACCTCCTGAGGC 256
DB 255 CAGAAGATACCTTTGTCCAAACACCATGCTCCATTACCTGGTTCTGCTGGAGTC 314
QY 257 CTGCTAGACTGTGCGGACATGTGGCGACTTTCTGGAGGGAACCCCAACCTCAGTC 316
DB 315 CCGTGGGAGTGCTCCAGGGCCATTACAGATTTTGGAGCGATACCCCATGTAATC 374
QY 317 TGAGATCTTACCGGGCCCTCTACTTCTGTGAGACCGCAAGGCTGAGCCGAGGGC 376
DB 375 TTTTATTATATAGACCGGCTTTATCACACGACAGATCCTCGAAATCGC---AAGGAC 431
QY 377 TCGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAAGATTATTTT 436
DB 432 TCAGGAGCTTATTAGCAGCGGTGTACTATCCAGATCATGACGGAAGAGTCTGGCT 491
QY 437 ACTGCTGGAATACTTTGTAGAAAAC 462
DB 492 ACTGCTGGAGGAATTTGTCAACTAC 517

RESULT 8
US-09-380-420C-1
Sequence 1, Application US/09380420C
Patent No. 6300544
GENERAL INFORMATION:
APPLICANT: Halkier, Barbara
Kahn, Rachel
Moller, Birger
TITLE OF INVENTION: Cytochrome P450 Monooxygenases
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Syngenta Patent Dept.
STREET: 3054 Cornwallis Road
CITY: RTP
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/380,420C
FILING DATE: 12-No. 6300544-1999
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-21251A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1929 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: P450ox
FEATURE:
NAME/KEY: CDS
LOCATION: 81..1673
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-380-420C-1

Query Match 6.4%; Score 38; DB 4; Length 1929;
Best Local Similarity 53.3%; Pred. No. 0.072;
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 236 CTGTGTTACCTCTGGAGCCCTGCTAGACTGTGCCGACATGTGGCGGACTTTCTGC 295
DB 793 CCAGCTTCTCCCGGAGGACTTCTCCCAACGCGCGCTCGCGGACCGCTCT 852
QY 296 GAGGAAACCCCACTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACC 355
DB 853 CGGGCTTCTCCCGCGCGGAGCGATCTTCAACGAGTCGACGCTCTTCTTCGAGAAG 912
QY 356 GCAAGCTGAGCCGAGGGGCTGCGGCGGC 385
DB 913 TCATGACCAACGACATGAGCCGCGCGCC 942

RESULT 9
US-09-189-060B-73
Sequence 73, Application US/09189060B
Patent No. 6270968
GENERAL INFORMATION:
APPLICANT: Dalboge, Henrik
Sandal, Thomas
APPLICANT: Kauppinen, Markus
APPLICANT: Borge, Diderichsen
TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences
FILE REFERENCE: 4772.204-US
CURRENT APPLICATION NUMBER: US/09/189,060B
PRIOR FILING DATE: 1998-11-10
PRIOR APPLICATION NUMBER: PCT/DK97/00216
PRIOR FILING DATE: 1997-05-12
NUMBER OF SEQ ID NOS: 74
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 73
LENGTH: 922
TYPE: DNA
ORGANISM: Hybrid
FEATURE:
NAME/KEY: CDS
LOCATION: (2)...(922)
US-09-189-060B-73
Query Match 6.3%; Score 37.8; DB 4; Length 922;

[illegible]

```

RESULT 14
US-09-015-188-1
; Sequence 1, Application US/09015188C
; Patent No. 639358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-015-188-1

Query Match          5.9%; Score 35.4; DB 4; Length 2190;
Best Local Similarity 48.7%; Pred. No. 0.47;
Matches 96; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 162 CTGCCACGTGGGAATTGCTCTCTCCGCTACATCTCGGACTCGGACCTAGACCCCTGGCCG 221
    || ||||| ||||| || ||||| || ||||| || ||||| || |||||
Db 343 CTCTTTGTTGGGCGACGCTCTTCAACCCAGCACCTGGAGCGTCTTCTACCTGTTTGAGCCCT 402
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 222 CTGCTACCGCGTCACTGTTTCACTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGT 281
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 CTACCACGTCCTGAGAACACGCTCATCCCGCGTCTCACCCAGGCAAGAGCCGCGGACCG 462
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 282 GGCCGCACTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTA 341
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 GGGGGTCATGCTAGGCGCCAGCGGACCTCTCTGGGAGCCCTCTACGACTGGCACCTCTA 522
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 342 CTCTCTGTGAGGACCGCA 358
    ||||| ||||| |||||
Db 523 CTCTCTGGAGAACTACA 539
    ||||| ||||| |||||

RESULT 15
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:06:39 ; Search time 80.0992 Seconds
(without alignments)
10937.110 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676
Perfect score: 597
Sequence: 1 atggacagccttctgatga.....ttctactttgggactttga 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues
Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	2818	9	US-09-966-880A-7
2	475.4	79.6	2440	9	US-09-966-880A-1
3	273.4	45.8	6564	9	US-09-966-880A-10
4	273.4	45.8	11204	9	US-09-966-880A-35
5	271	45.4	271	9	US-09-966-880A-13
6	148	24.8	148	9	US-09-966-880A-12
7	118.4	19.8	1534	10	US-09-729-674-173
8	114.8	19.2	987	10	US-09-925-300-699
9	114.8	19.2	1143	9	US-10-098-841-268
10	114.4	19.2	116	9	US-09-966-880A-14
11	110.4	18.5	1348	10	US-09-880-107-3282
12	104.4	17.5	2151	10	US-09-822-830A-359
13	78.4	13.1	371	9	US-09-918-995-30237
14	71.6	12.0	1404	9	US-10-198-846-13472
15	70.2	11.8	390	9	US-09-796-692-9051
16	70.2	11.8	390	9	US-10-040-862-9051
17	67	11.2	572	10	US-09-864-761-9416
18	66.6	11.2	328	9	US-09-796-692-5737
19	66.6	11.2	328	9	US-10-040-862-5737

20	63	10.6	476	9	US-09-918-995-32385	Sequence 32385, A
21	60	10.1	879	9	US-10-157-031-13	Sequence 13, Appl
22	54	9.0	2172	9	US-09-966-880A-15	Sequence 15, Appl
23	49.4	8.3	556	9	US-09-796-692-5061	Sequence 5061, Ap
24	49.4	8.3	556	9	US-10-040-862-5061	Sequence 5061, Ap
25	48	8.0	553	9	US-09-796-692-4504	Sequence 4504, Ap
26	48	8.0	553	9	US-10-040-862-4504	Sequence 4504, Ap
27	43.8	7.3	466	9	US-09-796-692-4007	Sequence 4007, Ap
28	43.8	7.3	466	9	US-10-040-862-4007	Sequence 4007, Ap
29	43.8	7.3	505	9	US-09-796-692-4443	Sequence 4443, Ap
30	43.8	7.3	505	9	US-10-040-862-4443	Sequence 4443, Ap
31	43.6	7.3	254	10	US-09-867-701-3576	Sequence 3576, Ap
32	43	7.2	997	10	US-09-864-761-9278	Sequence 9278, Ap
33	40.8	6.8	154	10	US-09-864-761-25835	Sequence 25835, A
34	38.4	6.4	261	9	US-10-015-219-763	Sequence 763, App
35	38.4	6.4	261	10	US-09-777-564-763	Sequence 763, App
36	38	6.4	1929	9	US-09-899-642-1	Sequence 1, Appl
37	37.2	6.2	473	9	US-10-046-938-1443	Sequence 1443, Ap
38	37.2	6.2	473	9	US-09-878-178-1443	Sequence 1443, Ap
39	37.2	6.2	473	9	US-10-146-502-1443	Sequence 1443, Ap
40	37	6.2	520	9	US-10-184-644-332	Sequence 332, App
41	37	6.2	520	9	US-10-184-634-332	Sequence 332, App
42	36.6	6.1	4689	9	US-09-860-846-34	Sequence 34, Appl
43	36.6	6.1	4689	9	US-09-988-384B-34	Sequence 34, Appl
44	36.6	6.1	4689	9	US-09-836-821-34	Sequence 34, Appl
45	36.6	6.1	4689	10	US-09-861-289-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Query Match 100.0%; Score 597; DB 9; Length 2818;
Best Local Similarity 100.0%; Pred. No. 6.1e-189;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGACAGCCTTCTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCCTCGG 60
|||||

```

Db 80 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 139
Qy 61 GCTAAGGGTCCGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db 140 GCTAAGGGTCCGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Qy 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGATGCTC 180
Db 200 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGATGCTC 259
Qy 181 TTCTCCGCTACATCTCGGACTCGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG 240
Db 260 TTCTCCGCTACATCTCGGACTCGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG 319
Qy 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 300
Db 320 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 379
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACCTGTGTGAGGACCGCAAG 360
Db 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACCTGTGTGAGGACCGCAAG 439
Qy 361 GCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGCTGCAATAGCCATCATGACC 420
Db 440 GCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGCTGCAATAGCCATCATGACC 499
Qy 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTTCAAA 480
Db 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTTCAAA 559
Qy 481 GCCTGGGAAGGGTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 560 GCCTGGGAAGGGTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 619
Qy 541 TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTTTGGGACTTTGA 597
Db 620 TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTTTGGGACTTTGA 676

```

```

RESULT 2
US-09-966-880A-1
; Sequence 1, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2440
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(686)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(92)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (690)...(2440)

```

FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(2440)
; OTHER INFORMATION: n = A,T,C or G
US-09-966-880A-1

```

```

Query Match          79.6%; Score 475.4; DB 9; Length 2440;
Best Local Similarity 87.3%; Pred. No. 2.8e-148;
Matches 521; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

```

Qy 1 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
Db 93 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 152
Qy 61 GCTAAGGGTCCGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db 153 GCCAAGGAGCGCATGAGACCTACCTGTGTAGTGAAGAGGAGATAGTGGCACC 212
Qy 121 TCCTTTTCACTGGAACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGATGCTC 180
Db 213 TCCTGTCTCACTGGACTTCGGCCACCTTCGCAACAAGTCTGGCTGCCACGTGGAATTTGG 272
Qy 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGCGTCACCTGG 240
Db 273 TTCTTACGCTACATCTCAGACTGGGACCTGGACCTGGCGGCGGCTGTACCGCGTCACCTGG 332
Qy 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTTCTCGGAGGG 300
Db 333 TTCACCTCTCGGAGCCCTGCTATGACTGTGCGCGGACGCTGGCTGTCTGAGATGG 392
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
Db 393 AACCCCAACCTCAGCTGAGGATTTTCACCGCGCGGCTCTACTTCTGTGAAGACCGCAAG 452
Qy 361 GCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGCTGCAATAGCCATCATGACC 420
Db 453 GCTGAGCCTGAGGGGCTGCGGAGACTGACCGCGGCTGGGGTCCAGATCGGGATCATGACC 512
Qy 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTTCAAA 480
Db 513 TTCAAAGACTATTTTACTGCTGGAATACATTTGTAGAAATCTGTAAGAACTTTCAAA 572
Qy 481 GCCTGGGAAGGGTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 573 GCCTGGGAAGGGTGCATGAAATTTCTGTCGCGGCTAACACAGACAACTTCGGCGCATCCTT 632
Qy 541 TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTTTGGGACTTTGA 597
Db 633 TTGCCCTCTGTATGAGGTTGATGACTTTCGAGATGCAATTTTCGTATGTTGGGATTTTGA 689

```

RESULT 3

```

US-09-966-880A-10
; Sequence 10, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10

```

; LENGTH: 6564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-10

Query Match 45.8%; Score 273.4; DB 9; Length 6564;
Best Local Similarity 99.6%; Pred. No. 2.1e-80;
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 155 AGAACGGCTGCCACGTGGAAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 214
|||||
Db 2590 AGAACGGCTGCCACGTGGAAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 2649
|||||
QY 215 CTGGCGCTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTCTAGGACTGTGCC 274
|||||
Db 2650 CTGGCGCTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTCTAGGACTGTGCC 2709
|||||
QY 275 GACATGTGCCGACTTCTTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 334
|||||
Db 2710 GACATGTGCCGACTTCTTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 2769
|||||
QY 335 GCCTCTACTTCTGTGAGGACCGAAGCTGAGCCCGGCTGCGGCGCTGCACCGCG 394
|||||
Db 2770 GCCTCTACTTCTGTGAGGACCGAAGCTGAGCCCGGCTGCGGCGCTGCACCGCG 2829
|||||
QY 395 CCGGGTGCAATAGCCATCATGACCTTCAAAGAT 429
|||||
Db 2830 CCGGGTGCAATAGCCATCATGACCTTCAAAGT 2864
|||||

RESULT 4

US-09-966-880A-35
; Sequence 35, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR FILING DATE: 2000-03-28
; PRIOR FILING DATE: 1999-12-27
; PRIOR FILING DATE: 1999-12-27
; PRIOR FILING DATE: 1999-06-24
; PRIOR FILING DATE: 1999-03-29
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35

LENGTH: 11204

TYPE: DNA

; ORGANISM: Homo sapiens
US-09-966-880A-35

Query Match 45.8%; Score 273.4; DB 9; Length 11204;
Best Local Similarity 99.6%; Pred. No. 2.9e-80;
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 155 AGAACGGCTGCCACGTGGAAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 214
|||||
Db 7805 AGAACGGCTGCCACGTGGAAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 7864
|||||
QY 215 CTGGCGCTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTCTAGGACTGTGCC 274
|||||
Db 7865 CTGGCGCTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTCTAGGACTGTGCC 7924
|||||
QY 275 GACATGTGCCGACTTCTTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 334
|||||
Db 7925 GACATGTGCCGACTTCTTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 7984
|||||

QY 335 GCCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGGCTGCGGCGCTGCACCGCG 394
|||||
Db 7985 GCCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGGCTGCGGCGCTGCACCGCG 8044
|||||
QY 395 CCGGGTGCAATAGCCATCATGACCTTCAAAGAT 429
|||||
Db 8045 CCGGGTGCAATAGCCATCATGACCTTCAAAGT 8079
|||||

RESULT 5

US-09-966-880A-13
; Sequence 13, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13

LENGTH: 271

TYPE: DNA

ORGANISM: Homo sapiens

US-09-966-880A-13

Query Match 45.4%; Score 271; DB 9; Length 271;
Best Local Similarity 100.0%; Pred. No. 2.2e-80;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AACGGCTGCCACGTGGAAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACCT 216
|||||
Db 1 AACGGCTGCCACGTGGAAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACCT 60
|||||
QY 217 GCGCGCTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTCTAGGACTGTGCCGA 276
|||||
Db 61 GCGCGCTGCTACCGCGTCACCTGGTTTCACTTCTGGAGCCCTCTAGGACTGTGCCGA 120
|||||
QY 277 CATGTGCCGACTTCTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGC 336
|||||
Db 121 CATGTGCCGACTTCTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGC 180
|||||
QY 337 CTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGGCTGCGGCGCTGCACCGCGC 396
|||||
Db 181 CTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGGCTGCGGCGCTGCACCGCGC 240
|||||
QY 397 GGGGTGCAATAGCCATCATGACCTTCAAAG 427
|||||
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAAG 271
|||||

RESULT 6

US-09-966-880A-12
; Sequence 12, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918

; PRIOR FILING DATE: 2000-03-28
 ; PRIOR APPLICATION NUMBER: JP 11-371382
 ; PRIOR FILING DATE: 1999-12-27
 ; PRIOR APPLICATION NUMBER: JP 11-178999
 ; PRIOR FILING DATE: 1999-06-24
 ; PRIOR APPLICATION NUMBER: JP 11-87192
 ; PRIOR FILING DATE: 1999-03-29
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 148
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-966-880A-12

Query Match 24.8%; Score 148; DB 9; Length 148;
 Best Local Similarity 100.0%; Pred. No. 2.3e-39;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 CCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGCTAAGGG 68
 Db 1 CCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGCTAAGGG 60
 QY 69 TCGCGGTGAGACTACTCTGCTAGCTAGTGAAGGCGTGACAGTGCTACATCTTTTC 128
 Db 61 TCGCGGTGAGACTACTCTGCTAGCTAGTGAAGGCGTGACAGTGCTACATCTTTTC 120
 QY 129 ACTGGACTTTGGTTATCTTCGCAATAAG 156
 Db 121 ACTGGACTTTGGTTATCTTCGCAATAAG 148

RESULT 7
 US-09-729-674-173
 ; Sequence 173; Application US/09729674
 ; Patent No. US20010039335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: LaVallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Steining, Robert J.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fectel, Kim
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
 ; FILE REFERENCE: 6055-64X
 ; CURRENT APPLICATION NUMBER: US/09/729,674
 ; CURRENT FILING DATE: 2000-12-04
 ; PRIOR APPLICATION NUMBER: 09/539,330
 ; PRIOR FILING DATE: 2000-03-30
 ; NUMBER OF SEQ ID NOS: 283
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 173
 ; LENGTH: 1534
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-729-674-173

Query Match 19.8%; Score 118.4; DB 10; Length 1534;
 Best Local Similarity 58.9%; Pred. No. 6.9e-29;
 Matches 225; Conservative 0; Mismatches 151; Indels 6; Gaps 1;
 QY 158 ACGCGTGCACGTGGAAATTCCTTCCTCCGTACATCTCGGAGGACCTAGACCTG 217
 Db 811 AAGGCGCCATGCAGAGCTGTCTCTCTGGAGCTGATTCCTTTTGGAGCTGGACCTGG 870

QY 218 GCGCTGCTACCGCTCACTGTTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGAC 277
 Db 871 ACCAGGACTACAGGTTACTGCTTCCCTCTCTGGAGCCCTGCTACGACTGTGCCAGG 930
 QY 278 ATGTGGCGGACTTTCTGCGAGGAAACCCCAACCTCAGTCTGAGGATCTTTCACGCGCGCC 337
 Db 931 AATGGCTAAATTCATTTTCAAAAAACAACACGTCGAGCTGTGCATCTTCACTGCGCGCA 990
 QY 338 TCTACTTCTGTGAGGACCGCAAGGCTGAGCCGCGGCGGTGCGGCGGTGACACCGCGCG 397
 Db 991 TCTA-----TGATGATCAAGGAAGATGTCAAGAGGGCTGCGCACCTTGGCCGAGGCTG 1044
 QY 398 GGTGCAATAGCCATCATGACCTTCAAAAGATTTATTTTACTGCTGGAATACTTTTGTAG 457
 Db 1045 GGGCCAAAATTTCAATAATGACATACAGTGAATTTAAGCACTGCTGGGACACCTTTTGGG 1104
 QY 458 AAAACCATGAAGAATCTTCAAGCCTCGGAGGCTGCATGAAATTTCACTGCTCTCT 517
 Db 1105 ACCACCGGATGTCCTTCCAGCCCTGGGATGAGTATGAGCACAGCAAGCACTGA 1164
 QY 518 CCAGACAGCTTCGCGCGCATCTCT 539
 Db 1165 GTGGAGGCTCGGCGCATCTCT 1186

RESULT 8
 US-09-925-300-699
 ; Sequence 699; Application US/09925300
 ; Patent No. US20020151681A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Craig Rosen,
 ; APPLICANT: Steve Ruben
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA101
 ; CURRENT APPLICATION NUMBER: US/09/925,300
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05988
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; NUMBER OF SEQ ID NOS: 1890
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 699
 ; LENGTH: 987
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-300-699
 Query Match 19.2%; Score 114.8; DB 10; Length 987;
 Best Local Similarity 53.9%; Pred. No. 8.7e-28;
 Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;
 QY 16 ATGAACCGGAGGAGTTTCTTTTACCAATTCAAAATGTCGCGTGGGCTAAGGTCGGGT 75
 Db 132 ATGTATCCAGGCACATTTCTACTTCCAATTTAAACACCTATGGAAGCAACGATCGGAAC 191
 QY 76 GAGACCTACCTGTGCTAGCTAGTGAAGGCGGTGACAGTGTACATCTCTTTCAGTGCAC 135
 Db 192 GAAACTTGGCTGTGCTTCAACCGTGAAGGTATAAGCCCGCTCAGTTGCTCTCTGGAG 251
 QY 136 TTTGGTTATCTTCGCA-----TAAGAACGGCTGCCACGTGGAATTCCTTTC 183
 Db 252 ACGGCGGTCTTCCGAAACCAAGTGGATTTCTGAGACCCATTTCTCATGAGAAAGGTGCTTC 311
 QY 184 CTCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCGCTGCTACCGGCTACCTGGTTC 243
 Db 312 CTCCTTGGTCTCGGAGACATCTGCTCTCTTAACAAAGTACCAGGTACCTGCTGTAC 371
 QY 244 ACCTCTGGAGCCCTGCTAGCACTGTCCCGACATGTGGCCGACTTTTCTCCGAGGAAAC 303
 Db 372 ACATCTTGGAGCCCTTCCCGACAGTGTGACGGGAGGTGGCCGAGTTCTTGGCCAGGCAC 431
 QY 304 CCCAACCTCAGTCTGAGGATCTTCCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCT 363

RESULT 9

RESULT 10

```
Query Match          19.2%; Score 114.4; DB 9; Length 116;
Best Local Similarity 99.1%; Pred. NO. 3.5e-28;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0
```

428 ATATATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAAGAACTTTTCAAGAGCTGGG 487
 1 ATTATTTTTTACTGCTGCAATACCTTTGTAGAAAACCATGAAAGAACTTTTCAAGAGCTGGG 60
 488 AAGGGCTGCATGAAATTTACGTTCTCTCCACACAGCTTCGGGCAATCTTTTG 543
 61 AAGGGCTGCATGAAATTTACGTTCTCTCCACACAGCTTCGGGCAATCTTTTG 116

RESULT 11
US-09-880-107-3282
; Sequence 3282, Application US/09880107
; Patent No. US2020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.

```

; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3282
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891
; NAME/KEY: unsure
; LOCATION: (1)..(1348)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3282

```

[illegible]

```

RESULT 12
US-09-822-830A-359
; Sequence 359, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulkota, Kamalakhar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES EN
; FILE REFERENCE: GIN 6402

```

```

: CURRENT APPLICATION NUMBER: US/09/822.830A
:
: CURRENT FILING DATE: 2001-03-29
:
: PRIOR APPLICATION NUMBER: 60/195,604
:
: PRIOR FILING DATE: 2000-04-06
:
: NUMBER OF SEQ ID NOS: 631
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 359
:
: LENGTH: 2151
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: US-09-822-830A-359

```

[illegible]

```

RESULT 13
US-09-918-995-30237
; Sequence 30237, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30237
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-30237

```

	Query Match	13.1%	Score 78.4;	DB 9;	Length 371;
	Best Local Similarity	64.1%;	Pred. No. 7.6e-16;		
	Matches 118; Conservative	0;	Mismatches 66;	Indels 0;	Gaps 0;
Qy	158	ACGGCTGCCACGTGGGAATTGCTTCTTCCTCGCTACATCTCGGACTGGGAGCTTAGACCCCTG	217		
Dd	171	AAGCGGCCCATGCAGAGCCTGCTTGCTCTGCACGTGATTTCCTTTTGGAACTCGACCCTGG	230		

Qy	218	GCCGCTGTACCGCGTCACTGGTTACCTCCTGGAGGCCCTGCTACGACTGTGCCCGAC	277
Db	231	ACAGGACTACAGGGTTACCTGCCTTCACCTCCTGGAGCCCCTGCTTCAGCTTGCCCAGG	290
Qy	278	ATGCGCGCGACTTTCTCGGAGGGGACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCC	337
Db	291	AATGGCTAAATTCAATTCAAAAAACAACACGTGAGCCTGTGCATCTTCACTGCCCGCA	350
Qy	338	TCTA 341	
Db	351	TCTA 354	
RESULT 14			
US-10-198-846-13472			
:	Sequence 13472, Application US/10198846		
:	Publication No. US20030099974N1		
:	GENERAL INFORMATION:		
:	APPLICANT: Lillie, James		
:	APPLICANT: Xu, Yongyao		
:	APPLICANT: Wang, Youzhen		
:	APPLICANT: Steinmann, Kathleen		
:	TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS		
:	TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND		
:	FILE REFERENCE: MRI-049		
:	CURRENT APPLICATION NUMBER: US/10/198,846		
:	CURRENT FILING DATE: 2002-07-18		
:	PRIOR APPLICATION NUMBER: 60/306,220		
:	PRIOR FILING DATE: 2001-07-18		
:	NUMBER OF SEQ ID NOS: 14084		
:	SOFTWARE: FastSeq for Windows Version 4.0		
:	SEQ ID NO 13472		
:	LENGTH: 1404		
:	TYPE: DNA		
:	ORGANISM: Homo sapiens		
:	FEATURE:		
:	NAME/KEY: misc_feature		
:	LOCATION: 1, 2, 3, 1379, 1402, 1403, 1404		
:	OTHER INFORMATION: n = A,T,C or G		
:	US-10-198-846-13472		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:09:48 ; Search time 24.5713 Seconds
(without alignments)
6475.074 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676

Perfect score: 1101
Sequence: 1 atggacagcctcttgatgaa.....ttcgactcttgaggactttga 597

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xlp
-Q-/cgn2_1/USPTO.spool/US09966880/runat_14062003_175523_10297/app_query.fasta_1.9493
-DB=A_Geneseq_101002 -QMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USFR=US09966880 -CGN_1_1_632 -runat_14062003_175523_10297 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NGS_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1086	98.6	198	21	AAB24198	Human activation-1
2	1008	91.6	198	21	AAB24197	Mouse activation-1
3	390	35.4	384	20	AA42383	Amino acid sequenc
4	390	35.4	384	22	AAU39075	Human secreted pro
5	390	35.4	384	23	ABBS5784	Human polypeptide
6	388	35.2	384	21	AA484437	Amino acid sequenc
7	369.5	33.6	190	19	AAW77092	Human RNA editing
8	369.5	33.6	190	21	AA111973	Human RNA editing
9	369.5	33.6	190	22	AAU38905	Human polypeptide
10	369.5	33.6	210	22	AAU23550	Novel human enzyme
11	369.5	33.6	221	22	AAU40691	Human polypeptide
12	369.5	33.6	222	21	AAU37061	Human prostate can
13	369.5	33.6	222	22	AAU23537	Novel human enzyme
14	336.5	30.6	268	23	AAE15256	Human RNA metaboli
15	323.5	29.4	272	22	AAU23799	Novel human enzyme
16	309.5	28.1	163	23	ABU97879	Human secretory po
17	274	24.9	224	23	AAE16630	Human AAD45360 pro
18	274	24.9	224	23	AAE15900	Human APOBEC2 prot
19	271	24.6	222	19	AAW52131	A novel human mRNA
20	262.5	23.8	367	22	AAO13149	Human polypeptide
21	208.5	18.9	261	22	AAU30167	Novel human secret
22	204.5	18.6	169	22	AAU30166	Novel human secret
23	203	18.4	113	21	AAU00666	Human secreted pro
24	199	18.1	236	15	AAU38705	Apo-B RNA editing
25	198	18.0	127	22	AAU95163	Human protein sequ
26	193	17.5	229	15	AAU58704	Apo-B RNA editing
27	103.5	9.4	94	21	AAU01186	Human secreted pro
28	102	9.3	720	22	ABG62508	Drosophila melanog
29	99.5	9.0	925	23	AAO14246	Human presenilin e
30	97.5	8.9	298	18	AAW43397	Human imidazole
31	97.5	8.9	1588	23	ABU09437	H. influenzae DXR
32	95	8.6	293	22	ABG14103	Novel human diagno
33	94.5	8.6	568	22	AAU66474	Murine cell growth
34	92.5	8.4	559	18	AAW43396	Human imidazole
35	92.5	8.4	2342	22	ABU60584	Drosophila melanog
36	90	8.2	242	22	AAU62666	Propionibacterium
C 37	90	8.7	581	20	AAU14543	Human lipolysis st
C 38	90	8.7	581	21	AAU95824	Human lipolysis st
C 39	90	8.7	581	22	AAU59903	Human leptin fragm
C 40	90	8.7	630	20	AAU14542	Human lipolysis st
C 41	90	8.7	630	21	AAU95823	Human lipolysis st
C 42	90	8.7	630	22	AAU59902	Human leptin fragm
C 43	90	8.7	649	20	AAU14541	Human lipolysis st
C 44	90	8.7	649	21	AAU95822	Human lipolysis st
C 45	90	8.7	649	22	AAU59901	Human leptin fragm

ALIGNMENTS*

RESULT 1

AAB24198
ID AAB24198 standard; Protein; 198 AA.

XX AAB24198;

XX 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase SEQ ID NO:8.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW immune related disease; allergy; allergic disease; anti-allergic;
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; IgG subclass selection disorder.
 XX Homo sapiens.
 XX WO200058480-A1.
 XX 05-OCT-2000.
 XX 28-MAR-2000; 2000WO-JP01918.
 XX 29-MAR-1999; 99JP-0087192.
 XX 24-JUN-1999; 99JP-0178999.
 XX 27-DEC-1999; 99JP-0371382.
 XX (NLSB) JAPAN TOBACCO INC.
 XX (HONJ/) HONJO T.
 XX Honjo T, Muramatsu M;
 XX WPI; 2000-611715/58.
 XX N-PSDB; AAC55312.
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX Claim 1; Page 140-141; 174pp; Japanese.
 XX The present sequence is human activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatologic activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX SQ Sequence 198 AA;

Alignment Scores:
 Pred. No.: 8.56e-115 Length: 198
 Score: 1086.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.64% Indels: 0
 DB: 21 Gaps: 0

US-09-966-880A-7_COPY_80_676 (1-597) x AAB24198 (1-198)

QY 1 ATGCAGACCTCTGTATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
 DB 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTyr 20
 QY 61 GCTAAGGTCGGCTGAGACCTACCTGCTAGCTAGTAGAGGCGGTGACAGTGTACA 120
 DB 21 AlaLysGlyArgArgGlnThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 QY 121 TCCTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGCTGCCACGTTGGGAATGCTC 180
 DB 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 QY 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACTGG 240
 DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

QY 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTTCTGCGAGGG 300
 DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
 DB 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 QY 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCCGCGGGTGCAGAAATAGCCATCATGACC 420
 DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 QY 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 480
 DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 QY 481 GCCTGGGAAGGCTGCATCAAAATTCAGTTCTGCTCTCCAGACACCTTCGGCGCATCTT 540
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 QY 541 TTGCCCTGTATGAGTTGATGACATTACGAGACGCAATTCGTACTTTGGGACTT 594
 DB 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2
 AAB24197
 ID AAB24197 standard; Protein; 198 AA.
 AC AAB24197;
 XX 05-FEB-2001 (first entry)
 DT Mouse activation-induced cytidine deaminase SEQ ID NO:2.
 DE Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; anti-allergic;
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder.
 XX Mus musculus.
 OS WO200058480-A1.
 PN 05-OCT-2000.
 XX 28-MAR-2000; 2000WO-JP01918.
 XX 29-MAR-1999; 99JP-0087192.
 XX 24-JUN-1999; 99JP-0178999.
 XX 27-DEC-1999; 99JP-0371382.
 XX (NLSB) JAPAN TOBACCO INC.
 XX (HONJ/) HONJO T.
 XX Honjo T, Muramatsu M;
 XX WPI; 2000-611715/58.
 XX N-PSDB; AAC55307.
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX Claim 1; Page 131-132; 174pp; Japanese.
 XX The present sequence is mouse activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and

CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
CC dermatologic activities, and can be used in gene therapy. AID
CC polynucleotides are useful in methods for identifying drugs for the
CC treatment of B cell associated immune system disorders, immunodeficiency
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
CC disease, DiGeorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC encoding AID may be used for gene therapy and the antibodies to the AID
CC protein may be used for diagnosis and treatment of these disorders.
XX
SQ Sequence 198 AA;

Alignment Scores:
Pred. No.: 6.61e-106 Length: 198
Score: 1008.00 Matches: 183
Percent Similarity: 95.94% Conservative: 6
Best Local Similarity: 92.89% Mismatches: 8
Query Match: 91.55% Indels: 0
DB: 21 Gaps: 0

US-09-966-880A-7_COPY_80_676 (1-597) x AAB24197 (1-198)

QY 1 ATGACAGCTCTTGATGAACGGAGGAGTTCTTTACCAATTCAAAATGCCGTGG 60
DB 1 MetAspSerLeuLeuMetLysGlnLysLysPheLeuTyrHisPheLysAsnValArgTrp 20
QY 61 GCTAAGGCGCGGTGAGACCTACCTGTCTAGTAGTGAAGGCGTGACAGTGTACA 120
DB 21 AlaLysGlyArgHisGlnThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40
QY 121 TCCTTTTACTGACTTGGTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 180
DB 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeuLeu 60
QY 181 TTCTCTCGCTACATCTCGACATCGGACCTAGACCTTGGCGCTGCTACCGCTCACTGG 240
DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTCGCGAGGG 300
DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
QY 301 AACCCCAACCTCAGCTGAGGATCTTCACGCGCGCTCTACTCTGTGTAGGACCGCAAG 360
DB 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 361 GCTGACCCGAGGGCTGCGGCTGCACCGCGCGGGTGCAGAAATAGCCATCATGACC 420
DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140
QY 421 TTCAAAGATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAACTTCAAA 480
DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160
QY 481 GCTGGGAGGGCTGCATGAATAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCTT 540
DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180
QY 541 TTGCCCCCTATGAGCTTGTATGACTTACGAGAGCGCATTTCTGCTACTTTGGGA 591
DB 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3
AAY42383
ID AAY42383 standard; Protein; 384 AA.
XX
AC AAY42383;
XX

DT 09-DEC-1999 (first entry)
XX Amino acid sequence of lp547_4.
DE
XX
KW secreted protein; cDNA library; clone; transmembrane protein;
KW signal sequence cloning; hybridization cloning; gene therapy;
KW receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 278..290
FT Protein /label= Leader/signal peptide
FT 291..384
FT /label= Mature protein
XX
PN WO9942470-A1.
XX
XX 26-AUG-1999.
XX
PF 18-FEB-1999; 99WO-US03458.
XX
PR 18-FEB-1998; 98US-0075038.
PR 17-FEB-1999; 99US-0251600.
XX
XX (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Merberg D;
PI Treacy M, Agostino MJ, Steininger RJ;
XX
DR WPI: 1999-518580/43.
DR N-PSDB; AA220856.
XX
XX New polynucleotides encoding human secreted proteins used for
PT therapeutic, diagnostic and research purposes.
PT
XX
PS Claim 17; Page 104-105; 125pp; English.
XX
XX This is the amino acid sequence of the lp547_4 protein, which is
CC derived from the lp547_4 clone isolated from a human adult blood cDNA
CC library.
CC The PNs and proteins of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC hematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity,
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumor invasion suppressor activity, and tumor
CC inhibition activity. The PNs are also stated to be useful for gene
CC therapy. Other activities include inhibiting the growth, infection or
CC function of bacteria, fungi, viruses and other parasites; effecting
CC bodily characteristics such as, e.g. weight, color, skin, etc.,
CC effecting biorhythms or cardiac cycles; enhancing fertility; treatment
CC of depression; treatment of pain; hormonal or endocrine activity.
XX
SQ Sequence 384 AA;

Alignment Scores:
Pred. No.: 2.22e-35 Length: 384
Score: 390.00 Matches: 83
Percent Similarity: 61.62% Conservative: 31
Best Local Similarity: 44.86% Mismatches: 59
Query Match: 35.42% Indels: 12
DB: 20 Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x AAY42383 (1-384)

QY 16 ATGAACCGAGGAGTTTCTTTACCAATTCAAAATGCCGTGGTTCGGCGGT 75
DB 197 MetAspProThrPheThrPheAsnAsnGluProTrpValArgGlyArgHis 216

QY 76 GAGACCTACTGTGCTACCTAGTGTGAGAGCGCTGACAGTGTACATCTTTTCACTGGAC 135
|||||
Db 217 GluThrTyrLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236
|||
QY 136 -----TTTGGTTATCTTCGCAATAG-----AACGGCTGC 165
|||||
Db 237 GluArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256
|||
QY 166 CACGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGACCTAGACCTGGCGCTGC 225
|||
Db 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspLeuAspGlnAsp 276
|||
QY 226 TACCGGCTCACCTGGTTACCTCCCTGGAGCCCTGCTACGACTGTGCTGGCCGCTGC 285
|||||
Db 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296
|||
QY 286 GACTTTCGCGAGGGAACCCAACTCAGTGTGAGATCTTCACCGCGCTCTACTTC 345
|||||
Db 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTyr--- 315
|||
QY 346 TGTGAGGACCGCAAGCTGAGCCGCGGCTGCGCGCTGCACCGCGCGGTGCAA 405
|||
Db 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334
|||
QY 406 ATAGCCATCATGACCTTCAAGATATTTTACTGCTGGATATCTTTGTAGAAAACCAT 465
|||||
Db 335 IleSerIleMetThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354
|||
QY 466 GAAAGAACTTCAAGCGCTGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAG 525
|||||
Db 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374
|||
QY 526 CTTGCGGCATCTT 540
|||||
Db 375 LeuArgAlaIleLeu 379
|||

RESULT 4

AAU39075
ID AAU39075 standard; Protein; 384 AA.

AC AAU39075;

DT 16-JAN-2002 (first entry)

DE Human secreted protein lp547_4.

KW Human; secreted protein; antiinflammatory; immunosuppressive;
KW neotropic; neuroprotective; antiarthritic; antimicrobial; vulnary;
KW cytosatic; antidiabetic; virucide; antifertility; anticonvulsant;
KW vasotropic; antiparkinsonian; immunostimulant; dermatological;
KW antihemmatic; antitumor; antiulcer; osteopathic; tranquiliser;
KW cerebroprotective; cytokine; cell proliferation; cell differentiation;
KW immune deficiency; severe combined immunodeficiency; SCID; tumour;
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
KW food supplement; vaccine.

OS Homo sapiens.

XX W0200175068-A2.

PN 11-OCT-2001.

PD 22-MAR-2001; 2001WO-US09369.

PF 30-MAR-2000; 2000US-0539330.

PR 04-DEC-2000; 2000US-0729674.

XX (GEMY) GENETICS INST INC.

PA

XX Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;
PI Clark H, Fechtel K, Merberg D;
XX WPI; 2001-639363/73.
DR N-PSDB; AAS59293.
XX Secreted human proteins, useful as vaccine for treating various
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and
PT nervous system disorders (e.g. stroke) -
XX Disclosure: Page 580-581; 619pp; English.
XX The invention relates to novel human secreted proteins, the nucleic
CC acids encoding them. The protein may exhibit cytokine, cell proliferation
CC or cell differentiation activity or may induce production of other
CC cytokines in certain cell populations and may exhibit immune stimulating
CC or immune suppressing activity, which is useful for the treatment of
CC various immune deficiencies and disorders e.g. severe combined
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation. The proteins are also useful in the treatment of diseases
CC and disorders including tissue, skin and organ transplantation and in
CC graft-versus-host diseases (GVHD). In the induction of tumour immunity,
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,
CC in the treatment of burns, incisions and ulcers; as well as in treatment
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by
CC inflammatory processes, diseases of the peripheral nervous system,
CC Alzheimer's, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,
CC infarction of cardiac and central nervous system vessel e.g. stroke,
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The
CC protein, having activin- or inhibin-related activities is useful as a
CC contraceptive based on the ability of inhibins to decrease fertility in
CC female mammals and decrease spermatogenesis in male mammals. The
CC proteins and nucleic acids are also useful as food supplements. The
CC present sequence represents a secreted protein of the invention.
XX
SQ Sequence 384 AA;

Alignment Scores:

Pred. No.: 2.22e-35 Length: 384
Score: 390.00 Matches: 83
Percent Similarity: 61.62% Conservative: 31
Best Local Similarity: 44.86% Mismatches: 59
Query Match: 35.42% Indels: 12
DB: 22 Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x AAU39075 (1-384)

QY 16 ATGAACGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGGCTAAGGCTCGCGGT 75
|||||
Db 197 MetAspProThrPheThrPheAsnPheAsnAsnGluProTrpValArgGlyArgHis 216
|||
QY 76 GAGACCTACTGTGCTACGTAGTAGAAGCGGTGACAGTGTCTATCTTTCACTGGAC 135
|||||
Db 217 GluThrTyrLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236
|||
QY 136 -----TTTGGTTATCTTCGCAATAG-----AACGGCTGC 165
|||||
Db 237 GluArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256
|||
QY 166 CACGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGC 225
|||
Db 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspLeuAspGlnAsp 276
|||
QY 226 TACCGGCTCACCTGGTTACCTCCCTGGAGCCCTGCTACGACTGTGCTGGCCGCTGC 285
|||||
Db 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296
|||
QY 286 GACTTTCGCGAGGGAACCCCACTCAGTGTGAGATCTTCACCGCGCGCTCTACTTC 345
|||||

Db 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTyr--- 315
QY 346 TGTGAGGACCGCAAGGCTGAGCCCGAGGGCGCTGCGCGCTGCACCGCGCGGGTGCAA 405
Db 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334
QY 406 ATAGCCATCATGCTTCAAGATATATTTTACTGTGGAATACTTTGTAGAAACCAT 465
Db 335 IleSerIleMetThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354
QY 466 GAAAGAACCTTCAAGACCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCACACAG 525
Db 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374
QY 526 CTTGCGCGCATCCTT 540
Db 375 LeuArgAlaIleLeu 379
RESULT 5
ID ABB55784
AC ABB55784; standard; Protein; 384 AA.
XX ABB55784;
XX
DT 14-FEB-2002 (first entry)
DE Human polypeptide SEQ ID NO 174.
XX
KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnery;
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory.
XX
OS Homo sapiens.
XX
PN US2001039335-A1.
XX
PD 08-NOV-2001.
XX
PF 04-DEC-2000; 2000US-0729674.
XX
PR 10-APR-1997; 97US-126425P.
PR 04-DEC-1997; 97US-067454P.
PR 02-DEC-1997; 97US-068379P.
PR 02-JAN-1998; 98US-070346P.
PR 07-JAN-1998; 98US-070643P.
PR 08-JAN-1998; 98US-070755P.
PR 13-JAN-1998; 98US-071304P.
PR 22-JAN-1998; 98US-072134P.
PR 30-JAN-1998; 98US-073095P.
PR 18-FEB-1998; 98US-075038P.
PR 30-MAR-2000; 2000US-0539330.
PR 23-NOV-1998; 98US-0197886.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (AGOS/) AGOSTINO M J.
PA (STEL/) STEININGER R J.
PA (SPAU/) SPAULDING V.
PA (WONG/) WONG G G.
PA (CLAR/) CLARK H. K.
PA (FECH/) FECHTEL K.
XX

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark H, Fechtel K;
XX
DR WPI: 2002-040725/05.
DR N-PSDB; ABA90962.
XX
PT New secreted proteins and encoding polynucleotides, useful in gene
PT therapies, particularly for preventing or treating autoimmune
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,
PT stroke or inflammations
XX
PS Disclosure; Page 311-312; 349pp; English.
XX
CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and
CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially
CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90985) and
CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in
CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
CC are deposited with the American Type Culture Collection (ATCC) with
CC accession number 98599. The polynucleotides and encoded polypeptides have
CC cytostatic, anti-inflammatory, immunomodulator, vulnerary,
CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic
CC and anti-inflammatory activity and acting as cytokine modulators,
CC haematopoiesis regulators, tissue growth modulators and/or cadherin
CC suppressors. The polypeptides and polynucleotides are useful in gene
CC therapies, particularly for preventing, treating or ameliorating any of
CC the following diseases: immune deficiency and disorders; e.g. bacterial
CC or fungal infections, autoimmune disorders, cancer, systemic lupus
CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell
CC deficiencies; wound, burn, incisions and ulcers, osteoporosis or
CC osteoarthritis; central and peripheral nervous system diseases and
CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
CC or systemic inflammatory response syndrome, ischaemia-reperfusion
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
CC foliaceus.
XX
SQ Sequence 384 AA;
Alignment Scores:
Pred. No.: 2,22e-35 Length: 384
Score: 390.00 Matches: 83
Percent Similarity: 61.62% Conservative: 31
Best Local Similarity: 44.86% Mismatches: 59
Query Match: 35.42% Indels: 12
DB: 23 Gaps: 3
US-09-966-880A-7_COPY_80_676 (1-597) x ABB55784 (1-384)
QY 16 ATGACCGGAGGAGTTTCTTACCAATTCAAAATGTCGCTGGGCTAGGGTCGGGT 75
Db 197 MetAspProThrPheThrPheAsnGluProTrpValArgGlyArgHis 216
QY 76 GAGACCTACCTGCTACGTAGTGAAGGCGTGCACAGTGTCTCTTCTCAGTGCAC 135
Db 217 GluThrTyrLeuCysTyrGluValGluArgMethHisAsnAspThrTrpValLeuLeuAsn 236
QY 136 -----TTTGGTTATCTTCGAATAAG-----AAGCGTCGC 165
Db 237 GlnArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256
QY 166 CAGGTGAATTGCTTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGC 225
Db 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspGlnAsp 276
QY 226 TACCGCTCACCTGGTTTCACTCTCTGGAGCCCTGCTACGACTGTGCCCATGTGGCC 285
Db 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296
QY 286 GACTTTCTGCGAGGAGACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCGCTCTACTTC 345

Db	17	PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys	36	PR	11-JUL-2000;	2000US-0217496.
QY	91	TACGTAGTAGAAG-----AGGCGTGACAGTGCTACATCCTTTCACTGGACTTGTGTAT	144	PR	14-JUL-2000;	2000US-0218290.
Db	37	PheThrValGluGlyIleLysArgSerValValSerTrpLysThr-----GlyVal	54	PR	26-JUL-2000;	2000US-0220963.
QY	145	CTTCGCAAT-----AAGAACGGCTGCCACGTGGAATTCCTTCCTCCGCTAC	192	PR	26-JUL-2000;	2000US-0220964.
Db	55	PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp	74	PR	14-AUG-2000;	2000US-0224518.
QY	193	ATCTCGGACTGGACCTAGACCTGCGCGCTGCTACCGGTGCTACCTGGTTCACCTCCTGG	252	PR	14-AUG-2000;	2000US-0224519.
Db	75	PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp	94	PR	14-AUG-2000;	2000US-0225213.
QY	253	AGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTCGGAGGAACCCCAACCTC	312	PR	14-AUG-2000;	2000US-0225214.
Db	95	SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal	114	PR	14-AUG-2000;	2000US-0225266.
QY	313	AGTCTGAGGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG	372	PR	14-AUG-2000;	2000US-0225267.
Db	115	AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu	133	PR	14-AUG-2000;	2000US-0225270.
QY	373	GGGCTCGGGGTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAAGATTAT	432	PR	14-AUG-2000;	2000US-0225271.
Db	134	GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe	153	PR	14-AUG-2000;	2000US-0225275.
QY	433	TTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTCAAAGCTGGGAAGG	492	PR	14-AUG-2000;	2000US-0225278.
Db	154	LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly	173	PR	14-AUG-2000;	2000US-0225279.
QY	493	CTGCATGAAATTCAGTTCGTCTCTCCAGACACTTCGGCGCATCCTT	540	PR	14-AUG-2000;	2000US-0225281.
Db	174	LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu	189	PR	14-AUG-2000;	2000US-0225282.
RESULT 10						
AAU23350						
ID	AAU23350	standard; Protein; 210 AA.				
AC						
XX	AAU23350;					
XX						
DT	17-DEC-2001	(first entry)				
XX						
DE		Novel human enzyme polypeptide #636.				
KW		Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;				
KW		ligase; hyperproliferative disorder; immunodeficiency disorder;				
KW		autoimmune disorder; neurological disorder; metabolic disorder;				
KW		inflammatory disorder; cardiovascular disorder; reproductive disorder;				
KW		blood-related disorder; infectious disorder; cytostatic; anti arthritic;				
KW		nephrotropic; anticoagulant.				
OS		Homo sapiens.				
XX						
PN	WO200153301-A2.					
XX						
PD	02-AUG-2001.					
XX						
PF	17-JAN-2001;	2001WO-US01239.				
XX						
PR	31-JAN-2000;	2000US-0179065.				
PR	04-FEB-2000;	2000US-0180628.				
PR	24-FEB-2000;	2000US-0184664.				
PR	02-MAR-2000;	2000US-0186350.				
PR	16-MAR-2000;	2000US-0189874.				
PR	17-MAR-2000;	2000US-0190076.				
PR	18-APR-2000;	2000US-0198123.				
PR	19-MAY-2000;	2000US-0205515.				
PR	07-JUN-2000;	2000US-0209467.				
PR	28-JUN-2000;	2000US-0214886.				
PR	30-JUN-2000;	2000US-0215135.				
PR	07-JUL-2000;	2000US-0216647.				
PR	07-JUL-2000;	2000US-0216880.				
PR	11-JUL-2000;	2000US-0217487.				

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
OS Homo sapiens.
XX WO200153312-A1.
PN 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Zhao J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI59847.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Example 2; SEQ ID NO 5622; 10078pp; English.
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX SQ Sequence 221 AA;

Alignment Scores:
Pred. No.: 3,93e-33 Length: 221
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.56% Indels: 9
DB: 22 Gaps: 4

US-09-966-880a-7_COPY_80_676 (1-597) x AAM40691 (1-221)

QY 31 TTCTCTTACCAATTCAAATGTCGGCTAAGGTCGGCGTGAGACCTACCTGTGC 90
Db 48 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 67
QY 91 TAGGTAGTGAAG-----AGGCGTGACAGTGCTACATCCCTTTCACTGGGACTTTGGTTAT 144
Db 68 PheThrValGluGlylleLysArgSerValValSerTrpLysThr-----GlyVal 85
QY 145 CTTCCGAAT-----AAGAACGGCTGCCAGCTGGAATTCCTTCCTCCGCTAC 192
DB: 11 Indels: 9 Gaps: 4

Db 86 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 105
QY 193 ATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTGCTACCTGCTTACCTCCCTGG 252
Db 106 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTrpThrSerTrp 125
QY 253 AGCCCTGTCTACGACTGTGCGCGACATGTGCGCGACTTCTCTCGAGGAGCAACCCACCTC 312
Db 126 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 145
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGTGAGGACCGCAAGGCTGAGCCCGAG 372
Db 146 AsnLeuThrIlePheThrAlaArgLeuTyrTrpPheGln---TyrProCysTyrGlnGlu 164
QY 373 GGGCTGGCGGCTGCACCGCGCGGTGCAATAGCATCATGACCTTCAAAAGCCTTGAAGGG 432
Db 165 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 184
QY 433 TTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTTCAAAAGCCTTGAAGGG 492
Db 185 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 204
QY 493 CTGATGAAATTCAGTTCGCTCTCCAGACAGCTTCCGGCGCATCCTT 540
Db 205 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 220

RESULT 12
AAB57061
ID AAB57061 standard; Protein; 222 AA.
XX AAB57061;
XX 13-MAR-2001 (first entry)
XX Human prostate cancer antigen protein sequence SEQ ID NO:1639.
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnar; gastrotentinal; nephrotropic; antinfetive; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX Homo sapiens.
XX WO200055174-A1.
PN 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05988.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM;
XX WPI; 2000-587513/55.
DR N-PSDB; AAF16264.
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX Claim 11; Page 2097-2098; 2338pp; English.
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnar, gastrotentinal,
CC nephrotropic, antinfetive, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen

CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.

XX
SQ Sequence 222 AA;

Alignment Scores:
Pred. No.: 3,93e-33 Length: 222
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.56% Indels: 9
DB: 21 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x AAB57061 (1-222)

QY 31 TTCTTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGCTGAGACCTACCTGTC 90
||| ::||| ||||| ::||| ||| ||||| ::||| |||
Db 49 PhetrPheGlnPheLysenLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuGly 68
:: ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 91 TAGCTAGTGAAG-----AGCGCTGACAGTGTACATCCTTTTCACTGGACTTGTGTTAT 144
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 69 PheThrValGluGlyLeuLysArgArgSerValValSerTrpLysThr-----GlyVal 86
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 145 CTTCGCNAAT-----AAGACGGCTGCCAGTGGAATGCTTCTCCCTGCTAC 192
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 87 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 106
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 193 ATCTCGACGTGGGACCTAGACCTGCGCTGCTACCGCTACCTGCTACCTGCTCCTGS 252
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 107 PheCysAspAspLeuSerProAsnThrLysTyrGlnValThrTrpTrpThrSerTrp 126
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 253 AGCCCTGTCTACGAGTGTGCCCGACATGTGCGCGACTTCTGCGAGGGAACCCCAACCTC 312
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 127 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 146
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG 372
::||| ||| ||||| ||||| ||||| :: ||| |||
Db 147 AsnLeuThrLlePheThrAlaArgLeuTyrPheGln---TyrProCysTyrGlnGlu 165
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 373 GGCTGCGGGGCTGCACCGCGCGGTCAATAGCCATCATGACCTTCAAGAGATTAT 432
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 166 GlyLeuArgSerLeuSerGlnGluGlyValAlaValAluIleMetAspTyrGluAspPhe 185
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 433 TTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTTCAAAGCCTGGGAGGG 492
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 186 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 205
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
QY 493 CTGCATGAAATTCAGTTCCTCTCCAGACACCTTCGGGGCATCCTT 540
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||
Db 206 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuGluSerLeu 221
||| ||| :: ||| :: ||| ||| ||||| :: ||| |||

RESULT 13

AAU23537

ID AAU23537 standard; Protein: 222 AA.

XX AAU23537;

XX AAU23537;

DT 17-DEC-2001 (first entry)

XX Novel human enzyme polypeptide #623.

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.

XX Homo sapiens.
OS WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0223270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0229284.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX

PA

(HUMA-) HUMAN GENOME SCI INC.

XX

PI

Rosen CA, Barash SC, Ruben SM;

XX

DR

WPI; 2001-465566/50.

DR

N-PSDB; AAS41407.

XX

PT

Novel polypeptides and polynucleotides useful for diagnosing,

PT

preventing, treating neural, immune system, muscular, reproductive,

PT

pulmonary, cardiovascular, renal, proliferative disorders and cancerous

PT

diseases

XX

PS

Claim 11; SEQ ID No 1533; 1180pp; English.

XX

CC

The present invention relates to the isolation of novel human enzyme

CC

polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences

CC

encoding them. The enzyme polypeptides of the invention may comprise the

CC

functional classes of oxidoreductases, transferases, hydrolases, lyases,

CC

isomerases or ligases. The sequences of the invention are useful in the

CC

diagnosis, treatment, prevention and/or prognosis of a wide range of

CC

disorders including hyperproliferative disorders (e.g. cancer),

CC

immunodeficiency disorders (e.g. AIDS) autoimmune disorders

CC

(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),

CC

metabolic disorders (e.g. phenylketonuria), inflammatory disorders

CC

(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),

CC

blood-related disorders (e.g. haemophilia), reproductive disorders

CC

(e.g. infertility) and infectious disorders (e.g. Influenza). The

CC

polynucleotides of the invention can also be used in gene therapy.

CC

AAU22915-AAU23814 represent the novel human enzyme polypeptides of the

CC

invention.

CC

Note: The sequence data for this patent did not form part of the printed

CC

specification, but was obtained in electronic format directly from WIPO

CC

at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ

Sequence 222 AA;

Alignment Scores:

Pred. No.: 3,93e-33 Length: 222

Score: 369.50 Matches: 79

Percent Similarity: 58.52% Conservative: 24

Best Local Similarity: 44.89% Mismatches: 64

Query Match: 33.56% Indels: 9

DB: 22 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x AAU23537 (1-222)

Qy 31 TTCTTTTACCAATTCAAAATGTCGGTGGCTAGGCTGCGGTGAGACCTACCTGTGC 90

Db 49 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 68

Qy 91 TAGCTAGTGAAG-----AGCGGTGACAGTCTACATCCTTTTCTTCTGCTGCTTTT 144

Db 69 PheThrValGluGlyIleLysArgSerValValSerIleLysThr-----GlyVal 86

Qy 145 CTTCGCAAT-----AAGACGGCTGCCACGTGGAATGTCTTCTTCGCTAC 192

Db 87 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 106

Qy 193 ATCTCGGACTGGGACCTAGACCTGGCGGTGCTACCGCTGCTACCTGTTACCTCCCTGG 252

Db 107 PheCysAspAlleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 126

Qy 253 AGCCCTGCTAGCACTGTCGGACATGTCGGACATTTCTTCGCGGAGGAGCCCAACCTC 312

Db 127 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 146

Qy 313 AGTCTGAGGATTTTCACTGGCGCTCTACTTCTGTGAGGACCGCAGGCTGAGCCGAG 372

Db 147 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 165

Qy 373 GGGCTGGCGGTGACCGCGCGGGGTGCAATAGCATCATGACCTTCAAGATTAT 432

Db 166 GlyLeuArgSerLeuSerGlnGlyValAlaValGluIleMetAspTyrGluAspPhe 185

XX OS Homo sapiens.
XX PN WO200155301-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01239.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 14, 2003, 18:12:33 ; Search time 17.1832 Seconds
(without alignments)
6680.033 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676
Perfect score: 1101
Sequence: 1 atggacagcctcttgatgaa.....ttcgtactttggactttga 597

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2.1/USPTO_spool/US09966880/runat_14062003_175525_10329/app_query.fasta_1.9493
-DB=PIR_73 -QFWT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 @CGN_1.1.456.@runat_14062003_175525_10329 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	
1	245	22.3	116	2	G01233	phorboln I - huma
2	225	20.4	236	2	A53853	apolipoprotein B m
3	216.5	19.7	236	2	I59323	apolipoprotein B m
4	211	19.2	229	2	JC4269	apolipoprotein B m
5	211	19.2	229	2	I48249	apolipoprotein B m
6	198	18.0	229	2	I59577	apolipoprotein B m
c 7	97.5	9.4	295	2	S58850	homeotic protein a
8	97.5	8.9	369	2	S63464	hypothetical prote
9	94.5	8.6	568	2	JC7317	cyclin-dependent k
c 10	94	9.1	504	2	JC1306	virion protein hom
11	93	8.4	277	2	A46241	interferon respons
12	92.5	8.4	2342	2	T13412	hypothetical prote
13	91.5	8.3	839	2	T04859	extensin homolog F
14	88.5	8.0	1493	2	A38218	GAP-associated pro

c 15	87.5	8.5	1791	2	T02345	hypothetical prote
16	86	7.8	522	2	S41819	nucleoporin p62 -
17	86	7.8	708	2	T19474	hypothetical prote
18	85.5	7.8	492	2	JC5169	alkaline nuclease
c 19	85	8.2	228	2	AD2935	hypothetical prote
c 20	85	8.2	1203	2	S26650	DNA-binding protei
c 21	84.5	8.2	521	2	D75581	malate synthase -
c 22	84.5	7.7	525	2	A35596	nuclear pore glyco
c 23	84.5	7.7	913	2	S20590	exo-alpha-sialidas
c 24	84	8.1	434	2	C75339	conserved hypoteth
c 25	83.5	7.6	383	1	VGBK6G	glycoprotein precu
26	82.5	7.5	481	2	T16484	hypothetical prote
27	82.5	7.5	492	2	T43064	CED-6 protein - Ca
c 28	81.5	7.4	219	2	I51382	achaete-scute homo
c 29	81.5	7.9	460	2	T33110	hypothetical prote
c 30	81.5	7.4	466	1	S15280	phosphogluconate d
c 31	81.5	7.9	800	2	E97935	phenylalanine-tRNA
c 32	81.5	7.4	1184	2	G01763	atrophin-1 - human
c 33	81.5	7.4	1184	2	S50832	atrophin-1 - human
c 34	81	7.8	235	2	A20688	Balbani ring 2 ch
c 35	81	7.8	513	2	A69831	probable acid-CoA
c 36	81	7.4	860	2	A96717	unknown protein, 4
37	80.5	7.3	184	2	I53641	mucin SAC - human
c 38	80.5	7.3	670	2	F36791	hypothetical prote
39	80.5	7.3	1446	1	A45344	immediate-early pr
c 40	80	7.7	218	2	G72754	probable phosphoen
c 41	80	7.3	476	2	C64601	fucosyltransferase
c 42	80	7.3	548	2	E70546	hypothetical prote
c 43	80	7.3	749	2	AD0345	probable phosphate
c 44	79.5	7.2	535	2	S65762	chitinase (EC 3.2.
c 45	79.5	7.7	654	2	T33044	hypothetical prote

ALIGNMENTS

RESULT 1
G01233
Chorbolin I - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 19-Apr-2002
C;Accession: G01233
R;Madsen, P.P.
submitted to the EMBL Data Library, December 1993
A;Reference number: G06330
A;Accession: G01233
A;Status: preliminary; translated from GB/EMBL/DBDJB
A;Molecule type: mRNA
A;Residues: 1-116 <MAD>
A;Cross-references: EMBL:U03891; NID:g436940; PIDN:AAA03706.1; PID:g436941
C;Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1

Alignment Scores:	2.31e-15	Length:	116
Pred. No.:	245.00	Matches:	51
Score:	61.82%	Conservative:	17
Percent Similarity:	46.36%	Mismatches:	32
Best Local Similarity:	22.25%	Indels:	10
Query Match:	2	Gaps:	3
DB:			

US-09-966-880A-7_COPY_80_676 (1-597) x G01233 (1-116)	
QY	226 TACCGCGTCACCGTTCACCTCTGAGGCCCTGCTTACGAC-----TGTGCCCGACAT 279
Db	7 TyrArgValThrTrpPheIleSerTrpSerProcysPheSerTrpGlyCysAlaGlyGlu 26
QY	280 GTGCGCGACCTTCTGCGAGGAAACCCAACTCAGTCTGAGGATCTTCACCGCGCGCTC 339
Db	27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgfile 46
QY	340 TACTTCTGTGAGGACCGCAGGCTGAGCCC-----CAGGGGCTGGCGGCGCTGCAC 390
Db	47 Tyr-----AspTyrAspProLeuTyrIysGluAlaLeuGlnMetLeuArg 61

391	CGCCCGGGTGCAAAATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAACTACT	450
QY		
Db		
62	ASPLAGLALGNALISLEIETHTRTYRSPGLUPHEGLUTYRCYSTRPASPTR	81
QY		
451	TTTTGTAGAAACCATGAAAGAACTTCAAGACCTGGGAAGGCTGCATGAAATTCAGTT	510
QY		
82	PhEValTYrArgGLInGLyCysPRoPhedInPRoTYrASpGLyLeuGLuGLuHisSerGLIn	101
Db		

RESULT 2

A53953
apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 18-Aug-2000
C:Accession: A53953
R:Yamanaka, S.; Poksay, T.K.S.; Balestra, M.E.; Zeng, G.Q.; Innerarity, T.L.
J. Biol. Chem. 269, 21725-21734, 1994
A:Title: Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A z
ributed.

Alignment Scores:		
Pred. No.:	1.99e-13	236
Score:	225.00	Matches: 49
Percent Similarity:	57.25%	Conservative: 26
Best Local Similarity:	37.40%	Mismatches: 48
Query Match:	20.44%	Indels: 8
DB:	2	Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x A53853 (1-236)

QY	70	CGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGTGACAGTGTGTACATCATCTTTTCA	129
Db	33	ArgLysGluAlaCysLeuLeuTyrGluLeuTyrGlyAlaSerSerLysThrTrpArg	52
QY	130	CTGAGCTTGGTTATCTCTCCGAATAAGAACGGTGC---CACGTGGAATGTCTCTTCCTC	186
Db	53	SerSer-----GlyLysAsnThrThrAsnHisValGluValAsnPheLeu	67
QY	187	---CGCTACATCTCGGACTCGGACCTAGACCTGGCGGTGTACCGCGTGCACCTGGTTC	243
Db	68	GlulysLeuThrSerGluGlyArgLeuGlyProSerThrCysSerIleThrTrpPhe	87
QY	244	ACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGACATTTCTGGCAGGGAAC	303
Db	88	LeuSerTrpSerProCysTrpGluCysSerMetAlaIleArgGluPheLeuSerGlnHis	107
QY	304	CCCAACCTCAGCTGTAGGATCTTACCGCGCGGCTCTACTTCTGTGAGACCGCAGGCT	363
Db	108	ProGlyValThrLeuIlePheValAlaArgLeuPheGlnHisMetAspArgArg---	126
QY	364	GAGCCGAGGGCTGCGCGGCTGTACCGCGCGGGTGCAATACCCATGACCTTC	423
Db	127	AsnArgGlnGlyLeuLysAspLeuValThrSerGlyValThrValArgValMetSerVal	146
QY	424	AAAGATTATTTTACTGCTGGAATACTTTGTA	456
Db	147	SerGluTyrCysTyrCysTrpGluAsnPheVal	157

RESULT 3

apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - human
C:Species: Homo sapiens (man)

C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 18-Aug-2000
C:Accession: I59323; S45253
R:Lau, P. P.; Zhu, H. J.; Baldini, A.; Charnsangavej, C.; Chan, L.
Proc. Natl. Acad. Sci. U.S.A. 91, 8522-8526, 1994
A:Title: Dimeric structure of a human apolipoprotein B mRNA editing protein
A:Reference number: I59323; M0ID:94355963; PMID:8078915
A:Accession: I59323
A>Status: translated from GB/EMBL/DDBJ

Alignment Scores:	
Pred. No.:	1.3e-12
Score:	216.50
Percent Similarity:	55.5%
Best Local Similarity:	34.9%
Query Match:	19.66%
DB:	2
Length:	236
Matches:	44
Conservative:	31
Mismatches:	42
Indels:	9
Gaps:	4

US-09-966-880A-7_COPY_80_676 (1-597) x R59323 (1-236)

103	QY	AGGCGTGACAGTGCTACATCTTTTCACTGACATTTGGTTATCTTCGC-----	150
		: : : : :	
33	Db	ArgLysGluAlaCysLeuLeuTyrGluIleLysTrpGlyMetSerArgLysIleTrpArg	52
151	QY	-----AATAAGACGGCTGC--CAGCTGGAATGCTCTTCCTC--CGCTACATCTCG	198
		: : : : :	
53	Db	SerSerGlyLysAsnThrThrAsnHisValGluValAsnPhelIleLysLysPheThrSer	72
199	QY	GACTGGGACCTAGACCTGGCGCGTGTACGGGTCACTGGTTCACTCTCGAGCCGC	258
		: : : : :	
73	Db	GluArgAspPheHisProSerIleSerCysSerIleThrTrpPheLeuSerTrpSerPro	92
259	QY	TGCTACGACTGTGCCGACATGTGGCGGACATTTCTGCGAGGGAAACCCCACTCA	318
		: : : : :	
93	Db	CysTrpGluCysSerGlnAlaIleArgGluPheLeuSerArgHisProGlyValThrLeu	112
319	QY	AGGATCTTCAACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTCAGCCCGAGGG	378
		: : : : :	
113	Db	ValIleTyrValAlaArgLeuPheTrpPheHisMetAspGlnGln---AsnArgGln	131
379	QY	CGGCGGTGTGACCCGCGCGGGGTGCAATAGCCATCATGACCTTCAAGATTATTTT	438
132	Db	ArgAspLeuValAsnSerGlyValThrIleGlnIleMetArgAlaSerGluTyrTrpHis	151
439	QY	TGCTGGAATACTTTTGTGA	456
152	Db	CysTrpArgAsnPheVal	157

RESULT 4

apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - mouse
C.Species: Mus musculus (house mouse)
C.Date: 10-Nov-1995 #sequence,revision 08-Feb-1996 #text,change 18-Aug-2000
JC4269

A; Accession: I49286

Alignment Scores:

Alignment Scores:

Pred. No.:	7, 82e-11	Length:	239
Score:	198.00	Matches:	56
Percent Similarity:	47.0%	Conservative:	24
Best Local Similarity:	32.94%	Mismatches:	62
Query Match:	17.98%	Indels:	28
DB:	2	Gaps:	5

US-09-966-880A-7_COPY_80_676 (1-597) x I59577 (1-229)

QY	70	CGCGCTGAGACCTACCTGTGCTACGTAGTAGAAG-----AGG	105
Db	33	ArgLysGluThrCysLeuLeuTyrGluIleAsnTrpGlyArgHisSerIleTrpArg	52
QY	106	CGTGACAGCTGCTACATCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGC	165
Db	53	HisThrSerGlnAsnThr-----AsnLys-----	60
QY	166	CACGTGAATTTGCTCTTCTCTC-----CGCTACATCTCGGACTGGACCTAGACCCCTGGCCGC	222
Db	61	HisValGluValAsnPheIleGluLysPheThrThrGluArgTyrPheCysProAsnThr	80
QY	223	TGCTACCGCGTGTGCTGGTCCACTCCTCGGAGCCCTGCTACGACTGTGCCGACATGTG	282
Db	81	ArgCysSerIleThrTrpPheLeuSerTrpSerProCysGlyLeuCysSerArgAlaIle	100
QY	283	GCCGACTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTAC	342
Db	101	ThrGluPheLeuSerArgTyrProHisValThrLeuPheIleTyrIleAlaArgLeuTyr	120
QY	343	TTCTGTGAGGACCGAAGCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTG	402
Db	121	HisHisAlaSerProArg---AsnArgGlnGlyLeuArgAspLeuIleSerSerGlyVal	139
QY	403	CAATAGCCATCATGACCTTCAAGATTATTTTTACTGCTGCAATACTTTTTGTAGAAAC	462
Db	140	ThrIleGlnIleMetThrGluGlnGluSerGlyTyrCysTrpArgAsnPheValAsnTyr	159
QY	463	CATGAAGAACATTTCAAACGCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGA	522
Db	160	SerProSerAsnGluAlaHisTrpProArgTyrProHisLeuTrpValArgLeu-----	177
QY	523	CACCTTCGGCGCATCTTTTGGCCCTGTRAT	552
Db	178	-----TyrValLeuGluLeuTyr	183

RESULT 7
homeotic protein abd-A - Junonia coenia
S58850
N:Alternate names: abdominal-A homeodomain protein
C:Species: Junonia coenia
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Nov-2000
R:Warren, R.W.; Nagy, L.; Selegue, J.; Gates, J.; Carroll, S.
C:Accession: S58850; S58851
Nature 372, 458-461, 1994
A:Title: Evolution of homeotic gene regulation and function in flies and butterflies.
A:Reference number: S58850; MUID:95075456; PMID:7840822
A:Accession: S58850
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-295 <WAR>
A:CROSS-references: EMBL:L41931; NID:G797276; PIDN:AAA68460.1; PID:G797277
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Accession: S58851
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 214-271 <WA2>
A:CROSS-references: EMBL:L42135; NID:G833750; PIDN:AAA68461.1; PID:G833751
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
C:Genetics:
A:Gene: abd-a
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;215-271/Domain: homeobox homology <HOX>

Alignment Scores:		
Pred. No.:	0.363	Length:
Score:	97.50	Matches:
Percent Similarity:	32.81%	Conservative:
Best Local Similarity:	21.88%	Mismatches:
Query Match:	9.43%	Indels:
DB:	2	Gaps:
		295
		42
		21
		58
		71
		7

US-09-966-880A-7_COPY_80_676 (1-597) x S58850 (1-295)

462 GTTTCCTACAAAGATATTCCAGCAGTAATAAATAATCTTTGAAGGTCAATGATGGCTATTGG 403
Db ||| :||| :||| :||| :||| :||| :||| :|||
25 ValHisTyrGlyAlaValProGlnGlnGlyHisGluMetGluGlyCysAspGlnGlnLeu 44
QY -----CCCCCGCGGTGCAGCCGCCGAC 376
Db ||| :||| :||| :||| :||| :||| :||| :|||
45 ArgProAlaGlnHisHisTyrProAlaGlnProAlaProGlyMetProTyrProArgphe 64
QY CCCCTCGGGCTCAGCCTTTGCGGTCTCTCACAGAAGTAGAGCGCCGCGTGAAGATCCCTCAG 316
Db ||| :||| :||| :||| :||| :||| :||| :|||
65 ProProTyrAspArgLeuGlyTyrTyrGlnGlnMetGluGlnAsnGlyTyrArgProAsp 84
QY ACTGAGGTGGGTTCCTCCGACAAAAGTCGGCCACATGTGGGC----- 271
Db ::| :||| :||| :||| :||| :||| :||| :|||
85 Ser-----ProSerGln---MetGlyHisMetGlyProLysThrAspGlyTyr 99
QY -----ACAGTC 265
Db ||| :||| :||| :||| :||| :||| :||| :|||
100 GlyProAsnGlyHisGlnProProThrProAlaValTyrThrSerCysLysLeuGlnAla 119
QY GTAGCAGGGCTCCAGGAGGTGAACCAGAGTGACGCGGTAGCAGCGCCAGGCTCAGGTC 205
Db ||||| :||| :||| :||| :||| :||| :||| :|||
120 AlaAlaGlyThrAlaGlyGlyValProGlySer----- 130
QY CCAGTCGCAGATGTAGCGGAGGAGAGACAATCCAGCTGGCAGCCGTCTTATTGCGAAG 145
Db ||| ::| :||| :||| :||| :||| :||| :||| :|||
131 ProProLeuGluGlnAlaGlnGlnMetProHisHis----- 142
QY ATAACCAAAGTCCAGTGAAGAGGATGTAGCACTGTCTACGCCCTTTCACGTACGTAGCAGACAG 85
Db :||| :||| :||| :||| :||| :||| :||| :|||
143 -----MethIspGlnHisMetAlaGln 151
QY -----ACCTTTAGCCCGCAGCATTTTTTCAATTGGTA 37
Db ||| :||| :||| :||| :||| :||| :||| :|||
152 HisGlyMetProProHisGlnGlnHisLeuMetTyrProValAspMetGlnHisGln 171
QY AAGANAATTCCTCGGTTCATCAAGAGCGTGCAT 1
Db :||| :||| :||| :||| :||| :||| :||| :|||
172 ThrGlnMetProProMetHisGlnGlnSerMethIs 183

RESULT 8
S63464
hypothetical protein YPL018w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein LpB13w
C:Species: Saccharomyces cerevisiae
C:Date: 16-May-1996 #sequence_revision 12-Jul-1996 #text_change 06-Feb-1998
C:Accession: S63464
R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;
submitted to the EMBL Data Library, September 1995
A:Reference number: S63452
A:Accession: S63464
A:Molecule type: DNA
A:Residues: 1-369 <WAN>
A:Cross-references: EMBL:U36624; NID:gl276642; PID:gl039459; MIPS:YPL018w
C:Genetics:
A:Gene: SGD:CTF19
A:Cross-references: SGD:S0005939; MIPS:YPL018w
A:Map position: 16L

Alignment Scores:
 Pred. No.: 0.366 Length: 369
 Score: 97.50 Matches: 50


```
Best Local Similarity: 27.50% Mismatches: 41
Query Match: 9.09% Indels: 65
DB: 2 Gaps: 7

US-09-966-880A-7_COPY_80_676 (1-597) x JC1306 (1-504)

QY 569 CGTAAGTCATCAACCTCATACAGGGGCAAA-----AGGATGCGC 531
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ArgSerGlySerThrArgThrArgGlyArgAlaAlaAlaArgSerThrThrGlyArgLeuGln 380
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 CGAAGCTGCTGGAGAGAGAACTGAATTTTCATGC-----AGCCCTTCCCGAGGCT 480
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 381 ArgProCysGlyProArgArgAlaAlaLysCysArgAlaThrProArgGlnArg 400
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 TTGAAGTCTCTTCATGGTTTTCTACAAAAGTATTCACGACGTAAATAATCTTTGAAG 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 401 LeuArg----- 402

QY 419 GTCATGATGGCTATTTGACACCGCGCGCGGTGCAGCGCGGCTCGGGCTCAGCC 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 -----AlaArgGlyGluProArgHisThrSerGlySerGly 414

QY 359 TTGCGGTCTCACAGAAGTAGAGCGCGGTGAAGATCCTCAGACTGAGGTGGGTTTC 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 AlaPheSerGlnGly-----ArgArgProGlyArgValCysArgLeu-GlyTrpAlaCy 432

QY 299 CCTCGCAGAAAGTCGGCCA-----CATGT 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 slysAlaArgSerGlyProAlaArgGlyGlyProGlyProSerProValArgSerGlyLe 452

QY 275 CGGGCACATCGTAGCAGGGCTCCAGAGAGGTGAACCGAGTGCAGCGGTAGCAGCGCCA 216
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 452 uGlyLeuSerArgAlaArgGlySerPro----- 461

QY 215 GGCTCTAGTCCAGTCCGAGTGTAGCGGAGGAGACGAATTCACGTGGCAGCGGT 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 462 -----GlyProGlyProAlaCys---GlyGlyProSerArgAlaArgGlyGlyArg 477

RESULT 11
A46241
Interferon response element-binding factor IREBF-2 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46241
R:Yan, C.; Tamm, I.
Proc. Natl. Acad. Sci. U.S.A. 89, 8859-8863, 1992
A:Title: Molecular cloning and characterization of additional factors that bind to the i
A:Reference number: A46241, MUID:93028366; PMID:1409578
A:Accession: A46241
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-277 <YAN>
A:Cross-references: GB:L13610; NID:g293674; PIDN:AAA39332.1; PID:g293675
A>Note: sequence extracted from NCBI backbone (NCBIN:115299, NCBIF:115300)

Alignment Scores:
Pred. No.: 0.98 Length: 277
Score: 93.00 Matches: 42
Percent Similarity: 33.11% Conservative: 7
Best Local Similarity: 28.38% Mismatches: 46
Query Match: 8.45% Indels: 53
DB: 2 Gaps: 6

US-09-966-880A-7_COPY_80_676 (1-597) x A46241 (1-277)

QY 131 TGGACTTTGGTTATCTTCCATATAAGACGGTGCACGTGGAAATGCTCTCTCCCGCT 190
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 TrpThrProTrpThrGlyArgCysSerThrAlaAlaSerThrGlyCysArgTrpArgAla 137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 ACATCTCGGACTGGACCTAGACCTCGCGCTGCTACCGCGTCACT----- 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 ThrAlaAlaArgProThrArgThrThrAlaAlaGlyAlaArgHisProValGlyThrAla 157
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 238 ----- 238
Db 158 AlaAlaAlaThrGlyGlyAlaAlaAlaLeuGlyAspAlaGlyAlaAlaAspLeuGlu 177
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 239 ---GGTTACCTCCCTCGA-----GCCCTGTACGACTGTGCCCCACATGTGGCGC 286
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 AlaGlyAlaGlyProGlyArgGluAlaAlaProAlaAlaThrAlaAlaProSerLeuGlyPro 197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 ACTTTCTGCGGAGGAACCC---CAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACT 343
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 AlaLeu-AlaArgAlaProAspGlnProSerProAspLeuProGluAspPro----- 215
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 344 TCTGTGAGGACCG----- 370
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 -----SerProSerProProArgSerProAspProAlaArgGlyGlnGlyProGlyAr 233
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 AGGGGCTGCGCGGCTGCACCG-----CGCGGGGTGCANA 406
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 gAspArgGluAlaAlaLeuArgProCysArgSerGluSerProSerLeuGlyArgGlyProAr 253
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 TAGCCATCATGACCTTCAAAGA 428
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 gAlaHisProSerLeuGlnLys 260

RESULT 12
T13412
Hypothetical protein I33E12.4 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13412
R:Murphy, L.; Harris, D.; Barrell, B.
Submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17668
A:Accession: T13412
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2342 <MUR>
A:Cross-references: EMBL:AL009192; NID:e1371524; PID:e1202150; PIDN:CAA15686.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0000667
A:Introns: 1161/3; 1205/1; 1283/1; 1432/3; 1489/2; 1912/3
A>Note: EG:I33E12.4

Alignment Scores:
Pred. No.: 1.19 Length: 2342
Score: 92.50 Matches: 37
Percent Similarity: 40.74% Conservative: 18
Best Local Similarity: 27.41% Mismatches: 41
Query Match: 8.40% Indels: 39
DB: 2 Gaps: 5

US-09-966-880A-7_COPY_80_676 (1-597) x T13412 (1-2342)

QY 173 AATTGCTCTTCCTCGCTACATCTCGACTGGACTAGACCTGGCCGCTGCTACCGCG 232
   ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1568 AsnValGlyAlaAspAlaAlaGlyAlaAlaGlyThrGlyThrValAlaAlaSerProGly 1587
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 233 TCACCTGGTTACCTCCCTGGAGCCCTGCTAGACTGTGCCGACATGTGGCGGACTTC 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1588 GlySerGlySer-----GlyAla-----Asp-IleSerAsnVal 1599
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 TGCAGAGGAACCCCAACCTCAGTCTGAG---ATCTTCACCGCGCGC----- 336
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1599 euArgGlyAsnProAsnIleSerMetArgGluLeuPheHisGlyGluGluLeuGly 1619
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 337 -----CTCTACTTCTGTGAGGACCGCGAGCTGAGCCGAGGGCTGCGCGC 382
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1619 aIGlnPheLysValProPheGlyCysSerSerGlnArgThrProGlyTrpThrA 1639
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 383 GCCTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAAGATTATTTTACTGCT 442
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1639 rgVal----- 1640
```

Plen. NO.:	2.63	Length:
Score:	88.50	Matches:

Flu. NO.:	2.63
Length:	1493
Score:	88.50
Matches:	28

Db 748 ThrArgSerArgThrProProArg---HisArgArgSerProSerValSerSerProGlu 766

```

QY 353 TCCTCACAAGTAGAGCGCGGTGAAGATCCTCAGACTGAGGTTGGGTTCCCTCGC 294
Db : : : : : ||| : : : ||| |||||
767 ProAlaGluLysSerArgSerArgArg-----ArgArgSerAlaSerSerProArg 784
QY 293 AGAAGTCGGCCACATGTCGGGCACACGTCG-----TAGCAGGGGCTCCAGGAG 246
Db ||| : : : : ||| : : : ||| |||||
785 ThrLysThrThrSerArgGlyArgSerProSerProLysProArgGlyLeuGlnArg 804
QY 246 ----- 246
Db
805 SerArgSerArgSerArgGluLysThrArgThrThrArgArgArgAspArgSerGly 824
QY 245 GTGAACCCAGGTGACGGGTAGCAGCGGCCAGGCTCTAGGTCCAGTCCGAGATGTAGCGG 186
Db : : : : : ||| : : : ||| ||||| : : : |||
825 SerSerGlnSerThrSerArgArgGlnArgSerArgSerArgSerArgValThrArg 844
QY 185 AGGAAGAGCAATTCCACAGTGGCAGCGTCTTATTCCGAAGATAACCAAGTCCAGTGAA 126
Db ||| ||| ||||| ||| : : : : |||||
845 Arg-Arg-----ArgGlyGlySerGlyTyr-----HisSerArgSerProAla-- 858
QY 125 AGGATGTAGCACTGTACCGCCTCTTCACTAGTACACAGGTAGGTCTCACGCCGACCC 66
Db ||| ||| ||||| ||| ||||| |||
859 -ArgGlnGluSerSerArgThrSerSerArgArgArgGlyArgGlySerArgThrProPr 878
QY 65 T 65
Db 878 O 878

```

Search completed: June 14, 2003, 18:56:35
Job time : 24.1832 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
c	1	364	33.1	382	1	PHB3_HUMAN	Q9uh17 homo sapien
	2	363.5	33.0	199	1	PHB1_HUMAN	P31941 homo sapien
	3	339	30.8	190	1	PHB2_HUMAN	Q9ue74 homo sapien
	4	225	20.4	236	1	ABME_RABIT	P47855 oryctolagus
	5	216.5	19.7	236	1	ABME_HUMAN	P41238 homo sapien
	6	211	19.2	229	1	ABME_MOUSE	P51908 mus musculus
	7	198	18.0	229	1	ABME_RAT	P38483 rattus norv
8	97.5	8.9	369	1	CT19_YEAST	Q02732 saccharomyc	
9	94	9.1	504	1	AT1N_HSVBP	P30020 bovine herp	
10	86.5	8.4	467	1	VE2_HPV24	P50770 human papil	
11	86	7.8	522	1	NU62_HUMAN	P37198 homo sapien	
12	85	7.7	503	1	MATK_KINBA	Q9tkc0 kunzea baxt	
13	85	7.7	525	1	SON_HUMAN	P18583 homo sapien	
14	84.5	7.7	525	1	NU62_RAT	P17955 rattus norv	
15	84	8.1	2404	1	SON_MOUSE	Q9qx47 mus musculus	
16	83.5	7.6	383	1	VGLZ_HSVK	P32515 equine herp	
17	81.5	7.4	1185	1	DRPL_HUMAN	P54259 homo sapien	
18	81	7.8	1719	1	DEND_RAT	P50617 rattus norv	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.


```
Best Local Similarity: 43.46% Mismatches: 57
Query Match: 33.02% Indels: 23
DB: 1 Gaps: 7

US-09-966-880A-7_COPY_80_676 (1-597) x PHB2_HUMAN (1-199)
QY 13 TTGTAGAACCGGAGGAAGTTTCTTACCAATTCAAAATTCACGCTGGCTAAGGCTCGG 72
Db 12 LeuMetAspProHisIlePheThrSerAsnPheAsnAsn-----GlyIleGlyArg 28
QY 73 CTGTAGACCTACCTGTGTACGTAGTGAAGAGCGTGACAGTGTACATCCTTTTCACGTG 132
Db 29 HisLysThrTyrLeuCystyrGluValGluArgLeuAspAsnGlyThrSerValLysMet 48
QY 133 GAC-----TTTGGTTATCTTCGCAAT-----AAGAAC-----GGC 162
Db 49 AspGlnHisArgGlyPheLeuHisAsnGlnAlaLysAsnLeuLeuLeuLeuLeuGlyPheTyrGly 68
QY 163 TGCCACGTGGAAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGC 222
Db 69 ArgHisAlaGluLeuArgPheLeuAspLeuValProSerLeuGlnLeuAspProAlaGln 88
QY 223 TCTACCGCGTACCTGTGTACCTCCTGAGCCCTGCTACGAC-----TGTGCCCGA 276
Db 89 IleTyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGly 108
QY 277 CATGTGGCGGACTTCTCGGAGGAGAACCCCACTCAGTCTGAGGATCTTCACCGCGCGC 336
Db 109 GluValArgAlaPheLeuGlnGluAsnThrHisValArgLeuArgIlePheAlaAlaArg 128
QY 337 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTGGCGCGGCTG 387
Db 129 IleTyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeu 143
QY 388 CACCGCGCGGGGTGCAATACCATCATGACCTTCAAGATATTTTACTGCTCGAAT 447
Db 144 ArgAspAlaGlyAlaGlnValSerIleMetThrTyrAspGlyPheLysHisCysTrpAsp 163
QY 448 ACTTTGTAGAAAACCATGAAAGAACTTTCAAGCGCTCGGAGGCGTGCATGAAATTC 507
Db 164 ThrPheValAspHisGlnGlyCysProPheGlnProTrpAspGlyLeuAspGluHisSer 183
QY 508 GTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 184 GlnAlaLeuSerGlyArgLeuArgAlaIleLeu 194

RESULT 3
PHB2_HUMAN
ID PHB2_HUMAN STANDARD; PRT; 190 AA.
AC Q9UE74;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phorbolin 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Keratinocytes;
RA Madsen P.;
RT "Molecular cloning of phorbolin 2.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC

or send an email to license@isb-sib.ch.
CC
CC EMBL: U61083; AAD00089.1; -.
CC InterPro: IPR002125; dCMP/cyt_deam.
CC PROSITE: PS00903; Cyt_DCMP_DEAMINASES; FALSE_NEG.
KW Hydrolase.
SQ SEQUENCE 190 AA; 22453 MW; A54DCBC100FC26F3 CRC64;

Alignment Scores:
Pred. No.: 9, 51e-23 Length: 190
Score: 339.00 Matches: 78
Percent Similarity: 55.26% Conservative: 27
Best Local Similarity: 41.05% Mismatches: 65
Query Match: 30.79% Indels: 20
DB: 1 Gaps: 6

US-09-966-880A-7_COPY_80_676 (1-597) x PHB2_HUMAN (1-190)
QY 16 ATGAACCGGAGGAAGTTTCTTACCAATTCAAAATTCGCGCTAAGGCTCGGCGT 75
Db 1 MetAspProAspThrPheThrPheAsnPheAsnAsnAspProLeuValLeuArgArgArg 20
QY 76 GAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTTTCACGTGAC 135
Db 21 GlnThrTyrLeuCystyrGluValGluArgLeuAspAsnGlyTrpMetValLeuMetAsp 40
QY 136 -----TTTGGTTATCTTCGCAAT-----AAGAAC-----GGCTGC 165
Db 41 GlnHisMetGlyPheLeuLeuGlnAlaLysAsnLeuLeuLeuLeuLeuGlyPheTyrGlyArg 60
QY 166 CACGTGGAATTTGCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGTGC 225
Db 61 HisAlaAspTrpArgPheLeuAspLeuValProSerLeuGlnLeuAspProAlaGlnIle 80
QY 226 TACCGCGTACCTGTGTACCTCCTGAGCCCTGCTACGAC-----TGTGCCGACAT 279
Db 81 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 100
QY 280 GTGCGCGACTTCTCGGAGGAGAACCCCACTCAGTCTGAGGATCTTCACCGCGCGCTC 339
Db 101 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 120
QY 340 TACTTCTGTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTGCGCGGCTGCAC 390
Db 121 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 135
QY 391 CGCGCGCGGTGCAATAGCATCATGACCTTCAAGATATTTTACTGCTGGAATACT 450
Db 136 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 155
QY 451 TTTGTAGAAAACCATGAAAGAACTTTCAAGCGCTGGGAGGCGTGCATGAAATTCAGTT 510
Db 156 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluHisSerGln 175
QY 511 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 176 AlaLeuSerGlyArgLeuArgAlaIleLeu 185

RESULT 4
ABME_RABIT
ID ABME_RABIT STANDARD; PRT; 236 AA.
AC P47855;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Apolipoprotein B mRNA editing protein (APOBEC-1).
GN APOBEC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
```

QY	70	CGCGCTGAGACCTACCTGTCTGTCTAGTCTAGAGAGCGGTGACAGCTGCTACATCCTTTTCA	129
Db	33	ArgLysGluAlaCysLeuLeuTy rGluIleLysTrpGlyAlaSerSerLysThrTrpArg	52
QY	130	CTGGAGCTTTGGTTATCTTCGCANAAGACCGGTGC---CACGTGGAAATGTCTCTCCCTC	186
Db	53	SerSer-----GlyLysAsnThrThrAsnHisValGluValAsnPheLeu	67
QY	187	---CGCTACATCTCGGACTGGAGCCTAGACCTCGCCGCTGCTACCCGCTACCTGGTTC	243
Db	68	GluLysLeuThrSerGluGlyArgLeuGlyProSerThrCysCysSerIleThrTrpPhe	87
QY	244	ACCTCCTGGAGCCCTGTCTAGACTGTGCCCGACATGTGCCACATTTCTGCGAGGAAAC	303
Db	88	LeuSerTrpSerProCysTrpGluCysSerMetAlaIleArgGluPheLeuSerGlnHis	107
QY	304	CCCAACCTCAGTCTGAGGATCTTCCGCGGGCCCTTACTTCTGTGAGGACCGCAAGGCT	363
Db	108	ProGlyValThrLeuIleIlePheValAlaArgLeuPheGlnHisMetAspArgArg---	126
QY	364	GAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAATACCCATCATGACCTTC	423
Db	127	AsnArgGlnGlyLeuLysAspLeuValThrSerGlyValThrValArgValMetSerVal	146

```

QY      424  AAAGATTATTTTACTGCTGGAACTACTTTTGA 456
Db      147  SerGluTyrCysTyrCysTrpGluAsnPheVal 157
          :::::  |||||
RESULT 5
ASWE_HUMAN
ID      ABME_HUMAN      STANDARD;      PRT;      236 AA.
AC      P41238;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Apolipoprotein B mRNA editing protein (APOBEC-1) (HEPR).
GN      APOBEC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eute-
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=942268910; PubMed=8208612;
RA      Hadjiagapiou C., Giannoni F., Funahashi T., Skarosi S.F.
RA      Davidson N.O.;
RT      "Molecular cloning of a human small intestinal apolipopro-
RT      editing protein.";
RL      Nucleic Acids Res. 22:1874-1879(1994).
RN      [2]
RP      SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX      TISSUE=Small intestine;
RA      Lau P.P., Zhu H.-J., Baldini A., Charnsangavej C., Chan H.
RT      "Dimeric structure of a human apolipoprotein B mRNA edit-
RT      ing and cloning and chromosomal localization of its gene.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:8522-8526(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      TISSUE=Peripheral blood leukocytes;
RA      Fujino T., Navaratnam N., Scott J.;
RT      "Human apolipoprotein B RNA editing deaminase gene (APOB-
RT      Genomics 47:266-275(1998).
CC      -1- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDIT-
CC      A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE ADE-
CC      -1- COFACTOR: ZINC (BY SIMILARITY).
CC      -1- SUBUNIT: HOMODIMER.
CC      -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE SMALL
CC      -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLA-
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through-
CC      between the Swiss Institute of Bioinformatics and the
CC      the European Bioinformatics Institute. There are no re-
CC      use by non-profit institutions as long as its content
CC      modified and this statement is not removed. Usage by a
CC      entities requires a license agreement (See http://www.istsb-
CC      or send an email to license@istsb-sib.ch).
CC      -----
DR      EMBL; L25877; AAA86766.1; -
DR      EMBL; L26234; AAA64230.1; -
DR      EMBL; AB009426; BAA23882.1; -
DR      EMBL; AB009422; BAA23882.1; JOINED.
DR      EMBL; AB009423; BAA23882.1; JOINED.
DR      EMBL; AB009424; BAA23882.1; JOINED.
DR      EMBL; AB009425; BAA23882.1; JOINED.
DR      PIR; S45253; S45253.
DR      Genew; HGNC:604; APOBEC1.
DR      MIT; 600130; -
DR      InterPro; IPR002125; dCMP/cyt_deam.
DR      Pfam; PF00383; dCMP_cyt_deam; 1.
DR      PROSITE; PS00093; CYT_DCMP_DEAMINASES; 1.
DR      mRNA processing; Hydrolase; Zinc.
DR      FT METAL 61 61
DR      ZINC (BY SIMILARITY).

```


Alignment Scores:

Pred. No.: 0.412 Length: 369
Score: 97.50 Matches: 50
Percent Similarity: 37.33% Conservative: 34
Best Local Similarity: 22.22% Mismatches: 70
Query Match: 8.86% Indels: 71
DB: 1 Gaps: 11

US-09-966-880A-7_COPY_80_676 (1-597) x CT19_YEAST (1-369)

QY 4 GACAGCCTTCTGATGACCGGAGGAGTTCTTTTACCAATTCAAAATGTCGCTGGCT 63
DB 62 AspAlaLeuThrArgArgAsnThrLeuLeuGlnGluLe 75
QY 64 AAGGGTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGGGGTGACAGTGTACATCC 123
DB 76 -----GlnThrTyrGlnAsnIleLeuMetLysGluAsnAsnSerLysThr--- 90
QY 124 TTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGTGCACCTGGAATTCGTC--- 180
DB 91 -----LysAsnGly-----AspIleLeuGln 97
QY 181 -----TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGC--- 219
DB 98 AsnAspIleThrGlnAspPheLeuAsnLeuIleSerIleSerSerSerAsnProAsnSer 117
QY 220 -----CGCTGCTACCGGCTCACCTGTTCACCTCCTGGAGC----- 255
DB 118 AlaIleSerAspArgLysArgValGluArgIleAsnGlyLeuThrAsnLeuGlnLysGlu 137
QY 256 -----CCTGCTACGACTGTGCCGACATGTGGCGACTTCTCGGAGGAACCCCAAC 309
DB 138 LeuValThrLysTyrAspThrLeuProLeuLeu-----Asn 149
QY 310 CTCAGTCGAGCATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAGGCTGAGCCC 369
DB 150 MetAsnLeuLeuSerTyrLeuArgAspHisThrTyrProHisLeuGlnValSerVal 169
QY 370 GAGGGCTCGCGGCTGCACCGCGGCTGCAATAGCCATCATGACCTTCAAGAT 429
DB 170 GlnSerArgAspArgValHisAsnAspGlyIleGluValLeuValAsnTyrLys--- 188
QY 430 TATTTTACTGCTGGAATACT-----TTGTAGAA 459
DB 189 -----PheCysArgAsnThrMetAsnProPheGluIleGlnPheLysMetPheTyrLys 206
QY 460 AACCATGAAGAACTTTCAAAGCCTGGGAAGGCTGCAT---GAAATTCAGTTCGTCTC 516
DB 207 PheGluAspSerThrLeuLeuLysTrpGluIleLeuArgIleSerThrAsnValArgLeu 226
QY 517 -----TCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTATGAG 555
DB 227 LysAlaLysGlnLeuLeuAlaThrArgAsnPheGlnLysCysLeuLeuSerLeuTyrGlu 246
QY 556 GTTGATGACTTACGA 570
DB 247 PheAspLysIleLys 251

RESULT 9

ATIN_HSVBP
ID ATIN_HSVBP STANDARD; PRT; 504 AA.
AC P30020;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Alpha trans-inducing protein (Alpha-TIF).
OS Bovine herpesvirus type 1 (strain p8-2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10324;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93012995; PubMed=1327963;

RA Carpenter D.E., Misra V.;
RT *Sequences of the bovine herpesvirus 1 homologue of herpes simplex
RL virus type-1 alpha-trans-inducing factor (UL48).";
Gene 119:259-263(1992).
CC -!- FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE
CC EARLY PROMOTERS (ALPHA GENES) (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11610; CAA77682.1; -.
DR PIR; S24229; S24229.
DR PIR; JCI306; JCI306.
DR InterPro; IPR003174; Alpha_TIF.
DR Pfam; PF02232; Alpha_TIF; 1.
KW Transcription regulation; Trans-acting factor; DNA-binding.
SQ SEQUENCE 504 AA; 54028 MW; 0031B4B0E31FCD25 CRC64;

Alignment Scores:

Pred. No.: 0.875 Length: 504
Score: 94.00 Matches: 44
Percent Similarity: 33.75% Conservative: 10
Best Local Similarity: 27.50% Mismatches: 41
Query Match: 9.09% Indels: 65
DB: 1 Gaps: 7

US-09-966-880A-7_COPY_80_676 (1-597) x ATIN_HSVBP (1-504)

QY 569 CGTAAGTCATCAACCTCATCAGGGGCAAA-----AGGATGCGC 531
DB 361 ArgSerGlySerThrArgThrArgGlyArgAlaAlaArgSerThrGlyArgLeuGln 380
QY 530 CGAAGCTGTCTGGAGAGACGAACGAATTTTCATGC-----AGCCCTCCCGAGCT 480
DB 381 ArgProCysGlyProArgArgArgAlaLysCysArgAlaThrProArgGlnArg 400
QY 479 TTGAAGACTTCTTTCATGGTTTCTTACAAAAGATTATCCAGCAGTAAAAATAATCTTTGAAG 420
DB 401 LeuArg----- 402
QY 419 GTCATGATGGCTATTTCACCCCGCGGGTGCAGCCCGCCGCGCTCAGCC 360
DB 403 -----AlaArgGlyGluProArgHisThrSerGlySerGly 414
QY 359 TTGGGTCCTCACAGAAGTAGAGCGCGGTGAAGATCCTCAGACTGAGGTGGGGTTC 300
DB 415 AlaPheSerGlnGly-----ArgArgProGlyArgValCysArgLeu-GlyTrpAlaCys 432
QY 299 CCTCGCAGAAAGTCGGCCA-----CATGT 276
DB 432 LysAlaArgSerGlyProAlaArgGlyGlyProGlyProSerProValArgSerGlyLe 452
QY 275 CGGGCAGACGCTAGCAGGGGCTCCAGAGGTGAACCAAGGTGACCGCGGTAGACGGCCA 216
DB 452 uGlyLeuSerArgAlaArgGlySerPro----- 461
QY 215 GGGTCTAGTCCAGTCCGAGATGTAGCGGAGGAGAGCAATTCACGTCGCGAGCCGT 158
DB 462 -----GlyProGlyProAlaCys---GlyGlyProSerArgAlaArgGlyArg 477
RESULT 10
VE2_HP24
ID VE2_HP24 STANDARD; PRT; 467 AA.
AC P50770;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)

ID	NUP62_HUMAN	STANDARD;	PRT; 522 AA.
AC	P37198; Q9NSL1; Q96C43;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Nuclear pore glycoprotein p62 (62 kDa nucleoporin).		
GN	NUP62.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92007939; PubMed=1915414;		
RA	Carmo-Fonseca M., Kern H., Hurt E.C.;		
RT	"Human nucleoporin p62 and the essential yeast nuclear pore protein		
RL	NSP1 show sequence homology and a similar domain organization.";		
RL	Eur. J. Cell Biol. 55:17-30(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RA	Bloeker H., Boecker M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDJB databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Skin, and Urinary bladder;		
RA	Strausberg R.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DDJB databases.		
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX.		
CC	THE N-TERMINAL IS PROBABLY INVOLVED IN NUCLEOTYOPASMIC		
CC	TRANSPORT. THE C-TERMINAL IS PROBABLY INVOLVED IN PROTEIN-PROTEIN		
CC	INTERACTION VIA COILED-COIL FORMATION AND MAY FUNCTION IN		
CC	ANCHORAGE OF P62 TO THE PORE COMPLEX.		
CC	-!- SUBCELLULAR LOCATION: CENTRAL REGION OF THE NUCLEAR PORE, WITHIN		
CC	THE TRANSPORTER. DURING MITOTIC CELL DIVISION, IT ASSOCIATES WITH		
CC	THE POLES OF THE MITOTIC SPINDLE.		
CC	-!- DOMAIN: CONTAINS F-X-F-G REPEATS.		
CC	-!- PTM: O-GLYCOSYLATED. CONTAINS ABOUT 10 N-ACETYLGLUCAMINE SIDE		
CC	CHAIN SITES PREDICTED FOR THE ENTIRE PROTEIN, AMONGST WHICH ONLY		
CC	ONE IN THE C-TERMINAL.		
CC	-!- SIMILARITY: BELONGS TO THE NUP62 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to licensel@sb-sib.ch).		
CC	-----		
DR	EMBL; X50521; CAAl1411.1; --		
DR	EMBL; AL162061; CAB82399.1; --		
DR	EMBL; BC003663; AAH03663.1; --		
DR	EMBL; BC014842; AAH14842.1; --		
DR	GlycoSuiteDB; P37198; --		
DR	GeneW; HGNC:8066; NUP62.		
DR	MTM; 605815; --		
KW	Nuclear protein; Transport; Glycoprotein; Coiled coil; Repeat;		
KW	Polymorphism.		
FT	DOMAIN 9 288		THR-RICH.
FT	DOMAIN 176 331		ALA-RICH.
FT	DOMAIN 1 295		15 X 9 AA APPROXIMATE REPEATS.
FT	REPEAT 1 9		1.
FT	REPEAT 13 21		2.
FT	REPEAT 29 37		3.
FT	REPEAT 39 47		4.
FT	REPEAT 57 65		5.
FT	REPEAT 71 79		6.
FT	REPEAT 85 93		7.
FT	REPEAT 111 119		8.
FT	REPEAT 137 145		9.
FT	REPEAT 155 163		10.
FT	REPEAT 168 176		11.


```
FT REPEAT 185 193 12.
FT REPEAT 234 242 13.
FT REPEAT 253 261 14.
FT REPEAT 287 295 15.
FT DOMAIN 328 458 COILED COIL (POTENTIAL).
FT CARBOHYD 373 373 O-LINKED (GLNAC) (BY SIMILARITY).
FT CARBOHYD 468 468 O-LINKED (GLNAC) (BY SIMILARITY).
FT VARIANT 233 233 A -> S (IN DBSNP:2290772).
/FTid=VAR_013467.
FT FT 283 283 T -> S (IN REF. 3: AAH14842).
FT CONFLICT 418 419 SG -> RA (IN REF. 1).
FT CONFLICT 431 431 E -> Q (IN REF. 1).
FT CONFLICT 507 507 E -> V (IN REF. 1).
SQ SEQUENCE 522 AA; 53268 MW; EAA3054F452719A2 CRC64;

Alignment Scores:
Pred. No.: 4.57 Length: 522
Score: 86.00 Matches: 56
Percent Similarity: 37.05% Conservative: 27
Best Local Similarity: 25.00% Mismatches: 89
Query Match: 7.81% Indels: 52
DB: 1 Gaps: 7

US-09-966-880A-7_COPY_80_676 (1-597) x NU62_HUMAN (1-522)
QY 35 TTTACCAATTCAAAATGTCCGCTGAGGCTGAGGCTGAGCCTACCTGTGCTACG 94
||||| : : : : : |||
Db 190 PheThrProAlaThrProAlaAlaThrAlaGlyAlaThrGlnProAlaAlaProThr 209
QY 95 TAGTGAAGAGCGGTGACAGTGCATACCTTTTCACGTGGAGCTTTGGTTACTTGC 151
||||| : : : : : |||
Db 210 ProThrAlaThrIleThrSerThrGlyPro-----SerLeuPheAlaSer 224
QY 152 ATAAGAGCGTCCGCGGTAATGCTCTCCCTCCGCTACATCTCGAGCTGGACCTAG 211
||||| : : : : : |||
Db 225 IleAlaThrAlaProThrSerSerAlaThrThrGlyLeuSerLeuCysThrProValThr 244
QY 212 ACCCTGCGCGCTGCTACCGCG-----TCACCTGGTTCACCT 247
||||| : : : : : |||
Db 245 ThrAlaGlyAlaProThrAlaGlyThrGlnGlyPheSerLeuLysAlaProGlyAlaAla 264
QY 248 CTGTGGAGCCCTGCTAGGACTGTGCCGACATGTGGCGGACTTTCTCGGAGGAAACCCCA 307
||||| : : : : : |||
Db 265 SerGlyThrSerThrThrThrSerSerAlaAlaThrAlaThrThrThrThrSer 284
QY 308 ACCTCAGCTGAGGATCTTCA-----CCGCGC 334
||||| : : : : : |||
Db 285 SerSerThrThrGlyPheAlaLeuAsnLeuLysProLeuAlaProAlaGlyIleProSer 304
QY 335 GCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGCGGCTGCGGCGTGCACCGCG 394
||||| : : : : : |||
Db 305 AsnThrAlaAlaAlaValThrAlaProGlyPro-GlyAlaAlaAlaGlyAlaAlaAl 324
QY 395 CCGGGTGCAAATAGCCATCATGACCTTCAAGATTATTTTACTGCTGAATCTTTG 454
||||| : : : : : |||
Db 324 aSerSerAlaMetThrTyrAlaGlnLeuGluSerLeu-IleAsnLysTrpSerLeuGluL 344
QY 455 TAGAAACCATGAAGAACTTTC-----AAAGCTGGGAA--- 489
||||| : : : : : |||
Db 344 euGluaspGlnGluArgHisPheLeuGlnGlnAlaThrGlnValAsnAlaTrpaspArg 364
QY 490 -----GGGCTGCATGAA----- 501
|||||
Db 364 hrLeuIleGluAsnGlyGluLysIleThrSerLeuHisArgGluValGluLysValLysL 384
QY 502 --AATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTTTCCCGCTGTATGAGTTG 559
||||| : : : : : |||
Db 384 euAspGlnLysArgLeuAspGlnGlnLeuAspPheIleLeuSerGlnGlnLysGluLeuG 404
QY 560 ATGACTTA 567
Db 404 luAspLeu 406
```

```
RESULT 12
MATK_KUNBA
* ID MATK_KUNBA STANDARD; PRT; 503 AA.
AC Q9PKC0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Kunzea baxteri (Scarlet kunzea).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Myrtales; Myrtaceae; Kunzea.
OX NCBI_TaxID=106042;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Brien M.M., Wilson P.G., Quinn C.J.;
RT "Molecular systematics of the Leptospermum suballiance (Myrtaceae).";
RL Aust. J. Bot. 48:621-628(2000).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC BY MITOCHONDRIAL INTRONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF104722; AAF05929.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 503 AA; 60203 MW; B114A4704FCEA059 CRC64;

Alignment Scores:
Pred. No.: 5.59 Length: 503
Score: 85.00 Matches: 35
Percent Similarity: 41.43% Conservative: 23
Best Local Similarity: 25.00% Mismatches: 52
Query Match: 7.72% Indels: 30
DB: 1 Gaps: 8

US-09-966-880A-7_COPY_80_676 (1-597) x MATK_KUNBA (1-503)
QY 91 TACGTAGTGAGAGCGGTGACAGTCTACATCTTTTCTACTCTGAC----- 135
||| : : : : : ||| : : : : : |||
Db 112 PheSerLeuGluArgLysGluIleAlaLysSerTyrAsnLeuArgSerIleHisSerIle 131
QY 136 TTTGGTTATCTTCGCAATAAG----- 156
||| : : : : : |||
Db 132 PheSerPheLeuGluAspLysPheThrHisLeuAspTyrValSerAspValLeuIlePro 151
QY 157 AACGGCTGCCAGTGGAAATGCTCTTC-----CTCCGCTACATCTCGGACCTGGGACCTA 210
||| : : : : : ||| : : : : : |||
Db 152 TyrHisIleHisLeuGluIleLeu***GlnThrLeuArgTyr-----TrpValLys 168
QY 211 GACCTTGGCCGCTGCTACCGGTACCTGGTTCACTCTCC-----TGGAGCCCTCTCTAC 264
||| : : : : : ||| : : : : : |||
Db 169 AspaLaserSerLeuHisLeuArgPhePheLeuHisAspTyrTrpAsnSerPheIle 188
QY 265 GACTGTGCCGACATGTGGCGACTTTCTGCGAGGGAACCCCAACTCAGTCTGAGGATC 324
||| : : : : : ||| : : : : : |||
Db 189 ThrProLysLysHisIleThrPhePheLeuLysGlyAsnProArgLeuPheLeu 208
QY 325 TTCACGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGGAGGGCTGCGGCGG 384
```

Db 209 TyrAsnSer-----HisIleCysGluTyrGluTyrIlePhePro---PheLeuArgAsn 225
Qy 385 -----CTGCACCGCGCGGGTGAATAATGACCATCATGACCTTCRAAGATTATTTTAC 438
Db 226 GlnSerSerHisLeuArgSerThrSerSerGlyIlePhe***GluArgIleTyrPheTyr 245

RESULT 13
ID SON_HUMAN
AC P18583; Q95981; Q9UPY0; Q14120; Q14487; Q9UKP9; Q9H7B1; Q9P070;
AC Q9P072;
DT 01-NOV-1990 (Rel. 16, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SON protein (SON3) (Negative regulatory element-binding protein) (NRE-
binding protein) (DBP-5) (Bax antagonist selected in saccharomyces 1)
DE (BASS1) (Protein C21orf50).
GN SON OR NREBP OR DBP5 OR C21ORF50 OR KIAA1019.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C; D; E AND F).
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedli M., Neergaard Henriksen C., Chapot F.,
RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
RA Antonarakis S.E.;
RT "From PRDs and open reading frames to cDNA isolation: revisiting the
RT human chromosome 21 transcription map.";
RL Genomics 78:46-54(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM G).
RC TISSUE=Liver;
RX MEDLINE=21316479; PubMed=11306577;
RA Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,
RA Ting L.-P.;
RT "Transcription repression of human hepatitis B virus genes by negative
RT regulatory element-binding protein/SON.";
RL J. Biol. Chem. 276:24059-24067(2001).
RN [3]
RP SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
RC TISSUE=Placenta;
RA Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,
RA Vitale L., Giannone S., Carinci P., Zannotti M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-130 FROM N.A.
RC TISSUE=Smooth muscle;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-114 FROM N.A.
RC TISSUE=Blood;
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";

RL DNA Res. 6:197-205(1999).
RN [7]
RP SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
RX MEDLINE=92049296; PubMed=1944255;
RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
RA Prasolov V.S.;
RT "Identification of a protein product of a novel human gene SON and
RT the biological effect upon administering a changed form of this gene
RT into mammalian cells.";
RL Mol. Biol. (Mosk) 25:731-740(1991).
RN [8]
RP SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
RC TISSUE=Placenta;
RX MEDLINE=93062885; PubMed=1435774;
RA Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
RT "The human son gene: the large and small transcripts contains various
RT 5'-terminal sequences.";
RL Mol. Biol. (Mosk) 26:807-812(1992).
RN [9]
RP SEQUENCE OF 1009-1131 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93062884; PubMed=1435773;
RA Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,
RA Chumakov I.M.;
RT "Coding part of the son gene small transcript contains four areas of
RT complete tandem repeats.";
RL Mol. Biol. (Mosk) 26:793-806(1992).
RN [10]
RP SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).
RX MEDLINE=93048367; PubMed=1424986;
RA Mattioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,
RA Lee J.S.;
RT "A cDNA clone for a novel nuclear protein with DNA binding
RT activity.";
RL Chromosoma 101:618-624(1992).
RN [11]
RP SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).
RX MEDLINE=89039788; PubMed=3054499;
RA Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;
RT "Decoding of the primary structure of the son3 region in human
RT genome: identification of a new protein with unusual structure and
RT homology with DNA-binding proteins.";
RL Mol. Biol. (Mosk) 22:794-801(1988).
RN [12]
RP SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).
RC TISSUE=Cerebellum;
RX MEDLINE=99439804; PubMed=10509013;
RA Greenhalf W., Lee J., Chaudhuri B.;
RT "A selection system for human apoptosis inhibitors using yeast.";
RL Yeast 15:1307-1321(1999).
RN [13]
RP FUNCTION: Represses hepatitis B virus (HBV) core promoter activity
RP and transcription of HBV genes and production of HBV virions.
CC Binds to the consensus DNA sequence: 5'-GA[GT]AN[CG][AG]CC-3'.
CC Might protect cells from apoptosis. Might be involved in pre-mRNA
CC splicing (By similarity).
CC -I- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.
CC -I- ALTERNATIVE PRODUCTS: 10 isoforms: A, B, C, D, E, F (shown here),
CC G, H, I and J; may be produced by alternative splicing.
CC -I- TISSUE SPECIFICITY: Widely expressed, with the higher expression
CC seen in leukocyte and heart.
CC -I- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC regions.
CC -I- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor
CC SFRS2/SC-35.
CC -I- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -I- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN
CC DUE TO A FRAMESHIFT.
CC -I- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN
CC DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC -!- FUNCTION: Transcriptional repressor. Binds to the consensus DNA sequence: 5'-GA[GT]AN[CG][AG]CC-3'. Might protect cells from apoptosis. Might be involved in pre-mRNA splicing (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are produced by alternative splicing.

CC -!- TISSUE SPECIFICITY: Widely expressed.

CC -!- DOMAIN: Contains 8 types of repeats which are distributed in 3 regions.

CC -!- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF193606; AAF23120.1; .

DR EMBL; AF193595; AAF23120.1; JOINED.

DR EMBL; AF193596; AAF23120.1; JOINED.

DR EMBL; AF193597; AAF23120.1; JOINED.

DR EMBL; AF193598; AAF23120.1; JOINED.

DR EMBL; AF193599; AAF23120.1; JOINED.

DR EMBL; AF193600; AAF23120.1; JOINED.

DR EMBL; AF193601; AAF23120.1; JOINED.

DR EMBL; AF193602; AAF23120.1; JOINED.

DR EMBL; AF193603; AAF23120.1; JOINED.

DR EMBL; AF193604; AAF23120.1; JOINED.

DR EMBL; AF193605; AAF23120.1; JOINED.

DR EMBL; AF193607; AAF23121.1; .

DR EMBL; AK019312; BAB31659.1; .

DR EMBL; AK019081; BAB31536.1; .

DR EMBL; AK008478; BAB25691.1; .

DR EMBL; AK008256; BAB25562.1; .

DR MGI; MGI:98353; Son.

DR InterPro: IPR001159; DS_RBD.

DR InterPro: IPR000467; G_patch.

DR Pfam; PF00035; dsrm; 1.

DR Pfam; PF01585; G_patch; 1.

DR SMART; SM00443; G_patch; 1.

DR PROSITE; PS0137; DS_RBD; 1.

DR PROSITE; PS0174; G_PATCH; 1.

KW RNA-binding; DNA-binding; Nuclear protein; Repeat;

KW Alternative splicing.

FT DOMAIN 721 850

FT 867 943

FT 1080 1080

FT 961 966

FT 961 966

FT 969 974

FT 976 981

FT 985 990

FT 993 998

FT 1001 1006

FT 1010 1015

FT 1018 1023

FT 1026 1031

FT 1035 1040

FT 1044 1049

FT 1055 1060

FT 1066 1071

FT 1075 1080

FT 1101 1133

FT 1910 1979

FT 1910 1916

FT 1938 1944

FT 1945 1951

FT 1952 1958

13 X 10 AA TANDEM REPEATS OF L-A-[ST]-[NSG]-[TS]-MDSQM.

11 X 7 AA TANDEM REPEATS OF [DR]-P-Y-R-[LI][AG][OHP].

14 X 6 AA REPEATS OF [ED]-R-S-M-M-S.

1-1.

1-2.

1-3.

1-4.

1-5.

1-6.

1-7.

1-8.

1-9.

1-10.

1-11.

1-12.

1-13.

1-14.

3 X 11 AA TANDEM REPEATS OF P-P-L-P-P-E-E-P-P-[TME]-[MTG].

7 X 7 AA REPEATS OF P-S-R-S-R-[TS].

2-1.

2-2.

2-3.

2-4.

FT REPEAT 1959 1965 2-5.

FT REPEAT 1966 1972 2-6.

FT REPEAT 1973 1979 2-7 (APPROXIMATE).

FT DOMAIN 1919 1990 2 X 19 AA REPEATS OF P-S-R-R-R-S-R-S-V-V-R-R-R-S-F-S-I-S.

FT REPEAT 1919 1937 3-1.

FT REPEAT 1980 1990 3-2 (APPROXIMATE).

FT DOMAIN 1991 2017 3 X TANDEM REPEATS OF [ST]-P-[VLI]-R-[RL]-[RK]-[RF]-S-R.

FT DOMAIN 2283 2329 G-PATCH.

FT DOMAIN 2349 2404 DRBM.

FT VARSPLIC 2086 2086 K -> F (IN ISOFORM 2).

FT VARSPLIC 2087 2404 MISSING (IN ISOFORM 2).

SQ SEQUENCE 2404 AA; 261428 MW; 648BF28ED3FC01D9 CRC64;

Alignment Scores:

Pred. No.: 8.06 Length: 2404

Score: 84.00 Matches: 59

Percent Similarity: 38.83% Conservative: 21

Best Local Similarity: 28.64% Mismatches: 72

Query Match: 8.12% Indels: 54

DB: 1 Gaps: 8

US-09-966-880a-7_COPY_80_676 (1-597) x SON_MOUSE (1-2404)

QY 581 CGAAATCGTCTCGTAAGTCATCAACCTCATACAGGGGCAAAAGGATCGCGGAAGCTGT 522

DB 1908 ArgThrProSerArgArgSerArgSerHisThrProSerArgArgArgSerIleSer 1927

QY 521 CTGGAGAGACGAAGTAAATTTTCATGCAGCCCTCCAGGCTTCCAGAGTGTGAAAGTCTTTTCATGG 462

DB 1928 ValGlyArgArgSerPheSerIleSerProSerArgArgSerArgThrProSerArg 1947

QY 461 TTTTCTACAAAAGTATTTCCAGCAGCAATAAATACTTTGAAGGTCATCATGCTATTGTC 402

DB 1948 ArgSerArgThrProSerArgArgSerArg----- 1957

QY 401 ACCCGCGCGGTGTCAGCGCG-----CGCAGCCCTCGGGCTCAGCC 360

DB 1958 ThrProSerArgArgSerArgThrProSerArgArgSerArgThrProSerArgArg 1977

QY 359 TTGCGGTCTCACAGAAAGTAGAGCGCGCGGTGAAGATCCTCAGACTGAGGTGGGTTC 300

DB 1978 ArgSerArgSerAlaValArgArgSerPheSerIleSerProValArgLeu----- 1995

QY 299 CCTCGCAGAAAGTCGCGCACATGTCGGCCACAGTCGTAGCAGGGGCTCCAGGAGGTGAAC 240

DB 1996 ---ArgArg-SerArgThrProLeuArgArgArgPheSerArgSerProIleArgArg 2014

QY 239 CAGGTGACGGGTAGCAGCGCGGCTAGGTCTAGGTCCAGTCCGCA-----GATGTAGCGG 186

DB 2014 sArgSerArg---SerSerGluArgGlyArgSerProLys-ArgLeuThrAspLeuAspL 2033

QY 185 AGGAAGACGAATTCACAGTG-----GCAGCGGTCTTATTTCGGAAGATAACCA 138

DB 2033 ysAlaGlnLeuLeuGluIleAlaLysAlaAsnAlaAlaMetCysAlaLys----- 2050

QY 137 AAGTCCAGTGAAAGGATGTAGCACTGTACAGCCCTCTTCACCTACGTACGACAGGTAGTTC 78

DB 2051 -----AlaGlyValProL 2055

QY 77 TCAGCGCGACCTTAGCCGCGGACATTTTTTGAATTGGTAAAGAAACTTCTCCGCGTTC 18

DB 2055 euProProAsnLeuLysProAlaPro----- 2067

QY 17 ATCAAGAGGTGTC 4

DB 2067 leGluGluLysVal 2071

Search completed: June 14, 2003, 18:29:36

Job time : 15.4884 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:12:03 ; Search time 36.6063 Seconds
(without alignments)
6720.711 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676

Perfect score: 1101

Sequence: 1 atggcagccttgatgaa.....ttcgtacttggacatttga 597

Scoring table: BLOSUM62

Ygapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09966880/runat_14062003_175524_10316/app_query.fasta_1.9493
-DB=SPTRMBL_21 -QFMT=fastan -SURFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 @CUN_1.1.978 @runat_14062003_175524_10316 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rviris.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	98.6	198	4 Q9GZX7	Q9gzx7 homo sapien

2	1008	91.6	198	11 Q9WVE0	Q9wve0 mus musculu
3	390	35.4	384	4 Q9HC16	Q9hc16 homo sapien
4	369.5	33.6	204	4 Q9Y555	Q9y555 homo sapien
5	367.5	33.4	190	4 Q96F12	Q96f12 homo sapien
6	354	32.2	294	4 Q9NRW3	Q9nrw3 homo sapien
7	349	31.7	429	11 Q99J72	Q99j72 mus musculu
8	336.5	30.6	386	4 Q96AK3	Q96ak3 homo sapien
9	315.5	28.7	139	4 Q9Y4V1	Q9y4v1 homo sapien
10	278	25.2	224	11 Q9W335	Q9wv35 mus musculu
11	274	24.9	224	4 Q9Y235	Q9y235 homo sapien
12	270.5	24.6	121	4 Q9Y533	Q9y533 homo sapien
13	257.5	23.4	103	4 Q9Y550	Q9y550 homo sapien
14	251	22.8	111	4 Q9UHI8	Q9uh18 homo sapien
15	240.5	21.8	121	4 Q9Y554	Q9y554 homo sapien
16	238.5	21.7	131	4 Q9Y930	Q9y930 homo sapien
17	238.5	21.7	245	4 Q9Y549	Q9y549 homo sapien
18	219	19.9	229	11 Q9EQP0	Q9eqp0 mesocricetu
19	218.5	19.8	236	4 Q9UM71	Q9um71 homo sapien
20	213.5	19.4	235	6 Q9TU17	Q9tui7 monodelphis
21	211	19.2	229	11 Q9U167	Q9u167 mus musculu
22	198	18.0	127	4 Q9H9H8	Q9h9h8 homo sapien
23	182.5	16.6	77	4 Q9Y551	Q9y551 homo sapien
24	179	16.3	132	4 Q9UE64	Q9ue64 homo sapien
25	102	9.3	720	5 Q9W293	Q9w293 drosophila
26	94.5	8.6	568	11 Q9EPR8	Q9epr8 mus musculu
27	93.5	8.5	568	11 Q9ESJ1	Q9esj1 mus musculu
28	93	8.4	277	11 Q06477	Q06477 mus musculu
29	92.5	8.4	2342	5 Q46048	Q46048 drosophila
30	91.5	8.3	839	10 Q9SN46	Q9sn46 arabidopsis
31	91	8.2	249	2 Q8VPM4	Q8vpm4 micrococcus
32	90	8.7	591	4 Q9UQL3	Q9uql3 homo sapien
33	89.5	8.1	629	12 Q9DWH3	Q9dwh3 rat cytomeg
34	89	8.6	429	4 Q00112	Q00112 homo sapien
35	88.5	8.0	1493	11 P81128	P81128 rattus norv
36	87.5	7.9	464	5 Q9NAX4	Q9nax4 dictyostell
37	87.5	8.5	583	4 Q9BT33	Q9bt33 homo sapien
38	87.5	8.5	585	4 Q9BWS2	Q9bws2 homo sapien
39	87.5	8.5	1275	4 Q9UQ36	Q9uq36 homo sapien
40	87.5	8.5	1783	4 Q15038	Q15038 homo sapien
41	87.5	8.5	1791	4 Q60382	Q60382 homo sapien
42	87.5	8.5	2296	4 Q9UHA8	Q9uha8 homo sapien
43	87.5	8.5	2752	4 Q9UQ35	Q9uq35 homo sapien
44	87	8.4	513	5 Q9VL71	Q9vl71 drosophila
45	86.5	7.9	209	10 Q9LWP8	Q9lwp8 oryza sativ

ALIGNMENTS

RESULT 1
Q9GZX7 PRELIMINARY; PRT; 198 AA.
ID Q9GZX7
AC Q9GZX7;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Activation-induced cytidine deaminase.
GN AID.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20408890; PubMed=10950930;
RA Muto T., Muramatsu M., Taniwaki M., Kinoshita K., Honjo T.;
RT "Isolation, tissue distribution and chromosomal localization of the
RL human activation-induced cytidine deaminase (hAID) gene.";
RN Genomics 68:85-88(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20460541; PubMed=11007475;
RA Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,
RA Catalan N., Forveille M., Dufourcq-Lageouse R., Gennery A.,

RA Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N.,
 RA Muramatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A.,
 RA Durandy A.,
 RT "Activation-induced cytidine deaminase (AID) deficiency causes the
 RT autosomal recessive form of the Hyper-IgM syndrome (HIGM2).";
 RL Cell 102:565-575(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-B-CELL;
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB040431; BAB12721.1; -;
 DR EMBL; AB040430; BAB12720.1; -;
 DR EMBL; BC006296; AAH06296.1; -;
 DR InterPro: IPR002125; dCMP/cyt_deam.
 DR PROSITE: PS00903; Cyt_DCMP_DEAMINASES; UNKNOWN.1.
 SQ SEQUENCE 198 AA; 23953 MW; 3C27BB143DB184A9 CRC64;

Alignment Scores:
 Pred. No.: 5.99e-101 Length: 198
 Score: 1086.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.64% Indels: 0
 DB: 4 Gaps: 0

US-09-966-880A-7_COPY_80_676 (1-597) x Q9GZX7 (1-198)

QY 1 ATGGACAGCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGCGCTGG 60
 DB 1 MetaspSerLeuLeuMetAsnArgLysPheLeuTyrglnPheLysAsnValArgTrp 20
 QY 61 GCTAAGGTCGGCGTGGAGACCTACCTGCTAGCTAGTGAAGGCGTGACAGTCTACA 120
 DB 21 AlaLysGlyArgArgGluThrTyrluLeuTyrluValValLysArgAspSerAlaThr 40
 QY 121 TCCTTTTCACTGGCTTGGTTATCTTCGCAATAAGACGGCTGCCAGCTTCTGCGAGT 180
 DB 41 SerPheSerLeuAspPheGlyTyrluLeuArgAsnLysAsnGlyCysHisValGluLeu 60
 QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGTGCTACCGGCTCACCTGG 240
 DB 61 PheLeuArgTyrlleSerAspTrpAspLeuAspProGlyArgCysTyrluValThrTrp 80
 QY 241 TTCACCTCTGGAGCCCTCTAGACTGTGCCGACATGTGCCGAGCTTCTGCGAGG 300
 DB 81 PheThrSerTrpSerProCysTyrluLeuArgHisValAlaAspPheLeuArgGly 100
 QY 301 AACCCCAACCTCAGTCTGAGGACTTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
 DB 101 AsnProAsnLeuSerLeuArgilePheThrAlaArgLeuTyrluPheCysGluAspArgLys 120
 QY 361 GCTGAGCCGAGGCGCTGCCGCGCTGCACCGCGCGGTGCAAAATAGCCATCATGCC 420
 DB 121 AlaGluProGluGlyLeuArgGluHisArgAlaGlyValGlnIleAlaIleMetThr 140
 QY 421 TTCAAAGATATTTTACTGCTGAATCTTTGTAGAAACCATGAAGAACTTTCAAA 480
 DB 141 PheLysAspTyrluPheTyrluAsnThrPheValGluAsnHisGluArgThrPheLys 160
 QY 481 GCCTGGAGGCGTGCATGAAATTCAGTTCGTCCTCCAGAGAGCTTGGCGCATCCTT 540
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGileLeu 180
 QY 541 TTCGCCCTGTATGAGTTGATGACTTACGAGAGCATTTTCGTAATTTGGGACTT 594
 DB 181 LeuProLeuTyrluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2

Q9WVE0

ID Q9WVE0

AC Q9WVE0; PRELIMINARY; PRT; 198 AA.

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Activation-induced cytidine deaminase.
 GN AICDA OR AID.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99303612; PubMed=10373455;
 RA Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,
 RA Davidson N.O., Honjo T.;
 RT "Specific expression of activation-induced cytidine deaminase (AID), a
 RT novel member of the RNA-editing deaminase family in germinal center B
 RT cells.";
 RL J. Biol. Chem. 274:18470-18476(1999).
 DR EMBL; AF132579; AAD41793.1; -;
 DR MGD; MGI:1342279; Aicda.
 DR InterPro: IPR002125; dCMP/cyt_deam.
 DR PROSITE: PS00903; Cyt_DCMP_DEAMINASES; UNKNOWN.1.
 SQ SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BEB2 CRC64;

Alignment Scores:
 Pred. No.: 4.42e-93 Length: 198
 Score: 1008.00 Matches: 183
 Percent Similarity: 95.94% Conservative: 6
 Best Local Similarity: 92.84% Mismatches: 8
 Query Match: 91.55% Indels: 0
 DB: 11 Gaps: 0

US-09-966-880A-7_COPY_80_676 (1-597) x Q9WVE0 (1-198)

QY 1 ATGGACAGCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGCGCTGG 60
 DB 1 MetaspSerLeuLeuMetLysGlnLysPheLeuTyrluHisPheLysAsnValArgTrp 20
 QY 61 GCTAAGGTCGGCGTGGAGACCTACCTGCTAGCTAGTGAAGGCGTGACAGTCTACA 120
 DB 21 AlaLysGlyArgHisGluThrTyrluLeuTyrluValValLysArgAspSerAlaThr 40
 QY 121 TCCTTTTCACTGGACTTGGTTATCTTCGCAATAAGACCGCTGCCAGCTTCTGCTC 180
 DB 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeu 60
 QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGTGCTACCGCGTCCACCTGG 240
 DB 61 PheLeuArgTyrlleSerAspTrpAspLeuAspProGlyArgCysTyrluValThrTrp 80
 QY 241 TTCACCTCTGGAGCCCTCTAGACTGTGCCGACATGTGCCGAGCTTCTTCTGCGAGG 300
 DB 81 PheThrSerTrpSerProCysTyrluAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
 QY 301 AACCCCAACCTCAGTCTGAGGACTTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
 DB 101 AsnProAsnLeuSerLeuArgilePheThrAlaArgLeuTyrluPheCysGluAspArgLys 120
 QY 361 GCTGAGCCGAGGCGTGGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGCC 420
 DB 121 AlaGluProGluGlyLeuArgGluHisArgAlaGlyValGlnIleGlyIleMetThr 140
 QY 421 TTCAAAGATATTTTACTGCTGAATCTTTGTAGAAACCATGAAGAACTTTCAAA 480
 DB 141 PheLysAspTyrluPheTyrluAsnThrPheValGluAsnArgGluArgThrPheLys 160
 QY 481 GCCTGGAGGCGTGCATGAAATTCAGTTCGTCCTCCAGAGAGCTTGGCGCATCCTT 540
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgileLeu 180
 QY 541 TTGCCCTGTATGAGTTGATGACTTACGAGAGCATTTTCGTAATTTGGGACTT 591
 DB 181 LeuProLeuTyrluValAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3		
Q9HC16	PRELIMINARY;	PRT; 384 AA.
ID	Q9HC16	
AC	Q9HC16;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	MDS019 (Phorbol-like protein MDS019).	
GN	MDS019.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RP	[1]	
RP	SEQUENCE FROM N. A.	
RP	Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.;	
RT	"Novel genes expressed in hematopoietic stem/progenitor cells from	
RT	Myelodysplastic Syndromes patient.";	
RT	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.	
LN	[2]	
RP	SEQUENCE FROM N. A.	
RP	TISSUE=SKIN;	
RA	Strausberg R.;	
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF182420; AAG14956.1; -.	
DR	EMBL; BC024268; AAH24268.1; -.	
DR	InterPro; IPR002125; dCMP/cyt deam.	
DR	PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN.1.	
SQ	SEQUENCE 384 AA; 46408 MW; 60525DC3B7D903D6 CRC64;	
Alignment Scores:		
Pred. No.:	1.09e-30	Length: 384
Score:	390.00	Matches: 83
Percent Similarity:	61.62%	Conservative: 31
Best Local Similarity:	44.86%	Mismatches: 59
Query Match:	35.42%	Indels: 12
DB:	4	Gaps: 3
US-09-366-880A-7_COPY_80_676 (1-597) x Q9HC16 (1-384)		
Qy	16 ATCAACGGAGGAGTTCITTACCAATTCAAAATGCTCGCTGGGTGAAGGTCGGCGT 75	:: ::
Db	197 MecAspProThrPheThrPheAsnPhaAsnGluProTrrpValArgGlyArgHis 216	:: ::
Qy	76 GAGACCTACCTGCTGCTACGTAGTGAAGAGGCGGTGACAGCTGCTACTCTCTTCTCACTGCAC 135	
Db	217 GluThrTyrluLeuTyrluValGluArgMethHisAsnAspThrTrpValLeuLeuAsn 236	
Qy	136 -----TTTGGTATCTTCGCAATAAG-----AACGGCTGC 165	
Db	237 GluArgArgGlyPheLeuLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256	
Qy	166 CAGTGGAAATGCTCTTCCTCCGCTACATCTCGGACCTGGGACCTAGACCTGGCGGCTGC 225	
Db	257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspGlnAsp 276	
Qy	226 TACCGCTCACCTGGTTCACCTCTGGAGGCCCTGCTACGACTGTCGCCGACATGTGCC 285	
Db	277 TyrArgValThrCysPheThrSerTrpSerProcysPheSerCysAlaGlnGluMetAla 296	
Qy	286 GACTTTCTCGAGGGAAACCCAACTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTTC 345	
Db	297 LysPheIleSerLysAsnLysHisValSerLeuLysIlePheThrAlaArgIleTyr--- 315	
Qy	346 TGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGC 405	
Db	316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334	
Qy	406 ATAGCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCAT 465	
Db	335 IlleSerIleMethThrTyrluSerGluPheLysHisCysTrpAspThrPheValIasPhisGln 354	
Qy	466 GAAAGAACTTTCAAAGCCTGGGAGGCGTGCATGAAAATTCAGTTTCGTCTCTCCAGACAG 525	

```
QY 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 LeuLysThrAsnPhleArgLeuLeuLysArgArgLeuArgGluSerLeu 203

RESULT 5
Q96F12 PRELIMINARY; PRT; 190 AA.
AC Q96F12;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=UTERUS;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC011739; AAF11739.1; -.
DR InterPro: IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 190 AA; 22827 MW; DA0584EF75C91CF0 CRC64;

Alignment Scores:
Pred. No.: 1,78e-28 Length: 190
Score: 367.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.38% Indels: 9
DB: 4 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x Q96F12 (1-190)
QY 31 TTCTTTTACCAATTCAAAATGTCCTGGCTAAGGCTGCGGTGAGACCTACCTGTGC 90
   ||| :|||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| |||
Db 17 PheTyPheGlnPhelLysAsnLeuTrpGluAlaAsnAspArgAspGluThrTrpLeuCys 36
   ||| :|||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| |||
QY 91 TAGCTAGTGAAG-----AGGCGTGACAGTGCTACATCTCTTTCACCTGGACTTGGTTAT 144
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 37 PheThrValGluGlyIleLysArgArgSerValSerTrpLysThr-----GlyVal 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 CTTTCGCAAT-----AAGAACGGCTGCACGTGGAATTGCTCTTCCTCCGCTAC 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 ATCTCGGACTGGACCTAGACCTGCGCGCTGCTACCGGTACCTGGTTACCTCCTG 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrglnValThrTrpTyThrSerTrp 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 253 AGCCCTGTACAGCTGTCCGCGCATGTGGCGGACTTCTCGGAGGGAACCCCAACCTC 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 313 AGTCTCAGAGCTTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAG 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyThrPheGln---TyrProCysTyrglnGlu 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 GGCTCGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTCAAAAGATTAT 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrglnAspPhe 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 433 TTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAAGCTTTCAAGCCCTGGGAAGG 492
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 LysTyrcysTrpGluAsnPheValTyrglnAsnAspAsnGluProPheLysProTrpLysGly 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 LeuLysThrAsnPhleArgLeuLeuLysArgArgLeuArgGluSerLeu 189
```

```
RESULT 6
Q9NRW3 PRELIMINARY; PRT; 294 AA.
AC Q9NRW3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Phorbollin I protein.
GN PBI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J., Huang Q., Yu Y., Xu S., Wang Y., Han Z., Chen Z., Zhou J.,
RA Tu Y., Gu W., Fu G., Huang C.;
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT Myelodysplastic Syndromes patient.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF165520; AAF86650.1; -.
DR InterPro: IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 294 AA; 33363 MW; 1B39C7A13D690901 CRC64;

Alignment Scores:
Pred. No.: 4,44e-27 Length: 294
Score: 354.00 Matches: 75
Percent Similarity: 59.06% Conservative: 26
Best Local Similarity: 43.86% Mismatches: 60
Query Match: 32.15% Indels: 10
DB: 4 Gaps: 5

US-09-966-880A-7_COPY_80_676 (1-597) x Q9NRW3 (1-294)
QY 31 TTCTTTTACCAATTCAAAATGTCCTGGCTAAGGCTGCGGTGAGACCTACCTGTGC 90
   ||| :|||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| |||
Db 17 PheTyPheGlnPhelLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
   ||| :|||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| |||
QY 91 TAGCTAGTGAAG-----AGGCGTGACAGTGCTACATCTCTTTCACCTGGACTTGGTTAT 144
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 37 PheThrValGluGlyIleLysArgArgSerValSerTrpLysThr-----GlyVal 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 CTTTCGCAAT-----AAGAACGGCTGCACGTGGAATTGCTCTTCCTCCGCTAC 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 ATCTCGGACTGGACCTAGACCTGCGCGCTGCTACCGGTACCTGGTTACCTCCTG 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrglnValThrTrpTyThrSerTrp 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 253 AGCCCTGTACAGCTGTCCGCGCATGTGGCGGACTTCTCGGAGGGAACCCCAACCTC 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 313 AGTCTCAGAGCTTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAG 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyThrPheGln---TyrProCysTyrglnGlu 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 GGCTCGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTCAAAAGATTAT 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrglnAspPhe 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 433 TTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAAGCTTTCAAGCCCTGGGAAGG 492
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 LysTyrcysTrpGluAsnPheValTyrglnAsnAspAsnGluProPheLysProTrpLysGly 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 Ile---LysAsnGlnLeuSerThrSerGluLys 183

RESULT 7
Q99J72
```


RA Peck A.;
RT "APOBEC-2, a cardiac and skeletal muscle specific member of the
RT cytidine deaminase supergene family.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=99333690; PubMed=10403781;
RA Liao W., Hong S.-H., Chang B.-J., Rudolph F.B., Clark S.C., Chan L.;
RT "APOBEC-2, a cardiac- and skeletal muscle-specific member of the
RT cytidine deaminase supergene family.";
RL Biochem. Biophys. Res. Commun. 260:398-404(1999).
DR EMBL; AL031778; CAB4740.1; -;
DR EMBL; AF161698; AAD45360.1; -;
KW Lipoprotein.
SQ SEQUENCE 224 AA; 25703 MW; CA0905AF8A8C8FA1 CRC64;

Alignment Scores:
Pred. No.: 4.96e-19 Length: 224
Score: 274.00 Matches: 64
Percent Similarity: 50.28% Conservative: 25
Best Local Similarity: 36.16% Mismatches: 76
Query Match: 24.89% Indels: 12
DB: 4 Gaps: 5

US-09-966-880A-7_COPY_80_676 (1-597) x Q9Y235 (1-224)

QY 31 TTTCTTTACCAATTCAAAATGTCGCTGGGTGAAGGTCGCGCTGAGACCTACCTGTGC 90
Db 52 PheLysPheGlnPheArgAsnValGluTyrSerGlyArgAsnLysThrPheLeuCys 71
QY 91 TACGTAGTGAAGAGCGGTGACAGTGCTACATCTTTTCACTGGACTTTGGTTATCTTCGC 150
Db 72 TyrValValGluAlaGlnGlyLysGlyGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91
QY 151 AATAAGAAC---GGCTGCACGTGGAATGCTCTTCCTCCGCTACATCTCGGACGTGGAC 207
Db 92 AspGluHisAlaAlaHisAlaGluAlaPhePheAsnThrIleLeuPro---Ala 110
QY 208 CTAGACCTGGCGGTGCTACCGCTACCTGCTCACCTGCTGAGCGCCCTGCTACGAC 267
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerSerProCysAlaAla 130
QY 268 TGTGCGCCGACATGTGGCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTC 327
Db 131 CysAlaAspArgIleIleLysThrLeuSerLysThrLysAsnLeuArgLeuLeuLeu 150
QY 328 ACCGGCGGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG-----GGGCTG 378
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluIleGlnAlaAlaLeu 166
QY 379 CGGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTAC 438
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186
QY 439 TGCTGGAATACTTTCTAGAAACACCATGAA-----AGAACTTTCAAGCCCTGGAA 489
Db 187 ValTrpGlnAsnPheValGluGlnGluGlyGluSerLysAlaPheGlnProTrpGlu 206
QY 490 GGGTGCATGAAATTCATGTCGTCTCCACAGACGCTTCGGCGCATCCTT 540
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIleLeu 223

RESULT 12
QY553 PRELIMINARY; PRT; 121 AA.
AC QY553;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BK150C2.10 (Putative novel Phorbol 1 like protein) (Fragment).
GN BK150C2.10.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45273.1; -;
FT NON_TER 1 1
FT NON_TER 121 121
SQ SEQUENCE 121 AA; 14309 MW; 97FC47DCDD82B247 CRC64;

Alignment Scores:
Pred. No.: 1e-18 Length: 121
Score: 270.50 Matches: 55
Percent Similarity: 60.00% Conservative: 17
Best Local Similarity: 45.83% Mismatches: 47
Query Match: 24.57% Indels: 1
DB: 4 Gaps: 1

US-09-966-880A-7_COPY_80_676 (1-597) x Q9Y553 (1-121)

QY 181 TTCCTCGGTACATCTCGGACTCGGACTAGACCTGCGCGCTGCTACCGGCTCACCTGG 240
Db 3 PheLeuSerTrpPheCysAspPheLeuSerProAsnThrAsnTyrGluValThrTrp 22
QY 241 TTCACCTCTGGAGCCCTGCTAGACTGTGCGCGACATGTGGCCGACTTTCGCGAGGG 300
Db 23 TyrThrSerTrpSerProGluCysAlaGlyGluValAlaGluPheLeuAlaArg 42
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 43 HisSerAsnValAsnLeuThrIlePheThrAlaArgLeuCysTyrPheTrpAspThrAsp 62
QY 361 GCTGACCCGAGGGGCTGGCGGCTGCACCGCGCGGGGTGCNAATAGCATCATGACC 420
Db 63 TyrGln---GluGlyLeuCysSerLeuSerGlnGlyAlaSerValLysIleMetGly 81
QY 421 TTCAAAGATTATTTTACTGCTGTAATCTTTGTAGAAAACCATGAAGACTTTCAAA 480
Db 82 TyrLysAspPheValSerCysTrpLysAsnPheValTyrSerAspGluProPheLys 101
QY 481 GCTGGGAAGGGTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 102 ProTrpLysGlyLeuGlnThrAsnPheArgLeuLeuLysArgArgLeuArgGluIleLeu 121

RESULT 13
QY550 PRELIMINARY; PRT; 103 AA.
AC QY550;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B
DE mRNA editing protein) and Phorbol) (Fragment).
GN BK150C2.6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45276.1; -;
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW Lipoprotein.
FT NON_TER 1 1
FT NON_TER 103 103
SQ SEQUENCE 103 AA; 12146 MW; 5DC969AE3ED348D9 CRC64;

Alignment Scores:

Job time : 40.6063 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 14, 2003, 18:13:03 ; Search time 5.75003 Seconds
(without alignments)
6109.704 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676
Perfect score: 1101
Sequence: 1 atggacagccttgatgaa.....ttcgtactttggactttga 597

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-Q-/cgn2_1/USPTO_spool/US09966880/runat_14062003_175525_10370/app_query.fasta_1.9493
-DB-issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -IOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 @cgn1.1.139 @runat_14062003_175525_10370 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCUTS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	33.6	190	1	US-08-816-241-1
2	369.5	33.6	190	3	US-09-128-395-1
3	271	24.6	222	1	US-08-687-895-1
4	271	24.6	222	2	US-09-040-482-1
5	245	22.3	116	1	US-08-687-895-3
6	245	22.3	116	2	US-08-816-241-3
7	245	22.3	116	2	US-09-040-482-3
8	245	22.3	116	3	US-09-128-395-3
9	216.5	19.7	236	1	US-08-687-895-4
10	216.5	19.7	236	1	US-08-816-241-4
11	216.5	19.7	236	2	US-09-040-482-4
12	216.5	19.7	236	3	US-09-128-395-4

13	210	19.1	236	1	US-08-158-682A-4	Sequence 4, Appli
14	198	18.0	229	1	US-08-158-682A-2	Sequence 2, Appli
15	198	18.0	229	1	US-08-015-203-2	Sequence 2, Appli
16	198	18.0	229	1	US-08-687-895-5	Sequence 5, Appli
17	198	18.0	229	1	US-08-816-241-5	Sequence 5, Appli
18	198	18.0	229	2	US-09-040-482-5	Sequence 5, Appli
19	198	18.0	229	3	US-09-128-395-5	Sequence 5, Appli
20	88.5	8.0	1513	5	PCT-US93-03076-2	Sequence 2, Appli
21	85.5	7.8	786	4	US-09-103-429A-3	Sequence 3, Appli
22	85.5	7.8	805	4	US-09-103-429A-4	Sequence 4, Appli
23	84.5	7.7	1128	4	US-09-627-650B-11	Sequence 11, Appli
24	84.5	7.7	1128	4	US-09-436-063C-11	Sequence 11, Appli
25	84.5	7.7	1652	4	US-09-627-650B-1	Sequence 1, Appli
26	84.5	7.7	1652	4	US-09-436-063C-1	Sequence 1, Appli
27	84.5	7.7	2508	4	US-09-627-650B-7	Sequence 7, Appli
28	84.5	7.7	2508	4	US-09-436-063C-7	Sequence 7, Appli
29	84.5	7.7	2544	4	US-09-627-650B-3	Sequence 3, Appli
30	84.5	7.7	2544	4	US-09-436-063C-3	Sequence 3, Appli
31	82.5	7.5	801	1	US-07-906-349A-6	Sequence 6, Appli
32	81.5	7.4	1185	4	US-09-041-886-23	Sequence 23, Appli
33	81	7.4	440	4	US-09-092-315-3	Sequence 3, Appli
34	81	7.4	575	3	US-08-922-865-2	Sequence 2, Appli
35	81	7.4	575	4	US-09-510-949-2	Sequence 2, Appli
36	81	7.4	1312	4	US-09-041-886-19	Sequence 19, Appli
37	80	7.3	267	4	US-08-818-112-142	Sequence 142, App
38	80	7.3	267	4	US-08-818-111-137	Sequence 137, App
39	80	7.3	267	4	US-09-056-556-142	Sequence 142, App
40	80	7.3	267	4	US-09-072-596-137	Sequence 137, App
41	80	7.3	476	4	US-09-092-315-5	Sequence 5, Appli
42	79.5	7.2	793	4	US-09-588-256-10	Sequence 10, Appli
43	79	7.6	135	4	US-09-419-485-1	Sequence 1, Appli
44	78.5	7.1	521	2	US-08-406-855A-19	Sequence 19, Appli
45	78.5	7.1	521	3	US-09-206-899-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-08-816-241-1
; Sequence 1, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surva K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-08-816-241-1

Alignment Scores:
Pred. No.: 3.67e-35 Length: 190
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.56% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-08-816-241-1 (1-190)
QY 31 TTCTTTACCAATTCAAAATGTCGCGTAAAGGTGCGGTGAGACCTACCTGTGC 90
Db 17 PhetYrPhelGlnPhelYsAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 91 TACGTAGTGAAG-----AGGCGTGACAGTGTACATCCTTTTCACTGGACTTGGTTAT 144
Db 37 PhetYrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 145 CTTCGCAAT-----AAGACGGCTGCCAGTGGAAATGCTTCTCTCCGCTAC 192
Db 55 PhetYrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 193 ATCTCGGACTGGACCTAGACCTGCGGTGCGGTGAGACCTACCTGTGC 90
Db 75 PhetYrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 253 ACCCGCTGTAGTCTTTGTAGAAACCATGAAAGACTTTCAAAGCGCTGGGAAGG 492
Db 95 SerProCysProAspCysAlaGlyValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGGAGCGCTGAG 372
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyTrpPheGln---TyrProCysTrpGlnGlu 133
QY 373 GGGCTGCGGCGCTGCACCGCGCGGTGCAAAATAGCCATGACCTTCAAAAGCTT 432
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTrpGluAspPhe 153
QY 433 TTTTACTGCTGAATCTTTGTAGAAACCATGAAAGACTTTCAAAGCGCTGGGAAGG 492
Db 154 LysTrpCysTrpGluAsnPheValTyTrpAsnAspAsnGluProPheLysProTrpLysGly 173
QY 493 CTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGGCATCCTT 540
Db 174 LeuLysThrAsnPheArgLeuLeuLysArgLeuArgLeuSerLeu 189

RESULT 2
US-09-128-395-1
; Sequence 1, Application US/09128395
; Patent No. 6087108
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surva K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-09-128-395-1

Alignment Scores:
Pred. No.: 3.67e-35 Length: 190
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.56% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-128-395-1 (1-190)
QY 31 TTCTTTACCAATTCAAAATGTCGCGTAAAGGTGCGGTGAGACCTACCTGTGC 90
Db 17 PhetYrPhelGlnPhelYsAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 91 TACGTAGTGAAG-----AGGCGTGACAGTGTACATCCTTTTCACTGGACTTGGTTAT 144
Db 37 PhetYrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 145 CTTCGCAAT-----AAGACGGCTGCCAGTGGAAATGCTTCTCTCCGCTAC 192
Db 55 PhetYrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 193 ATCTCGGACTGGACCTAGACCTGCGGTGCGGTGAGACCTACCTGTGC 90
Db 75 PhetYrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 253 ACCCGCTGTAGTCTTTGTAGAAACCATGAAAGACTTTCAAAGCGCTGGGAAGG 492
Db 95 SerProCysProAspCysAlaGlyValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGGAGCGCTGAG 372
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyTrpPheGln---TyrProCysTrpGlnGlu 133
QY 373 GGGCTGCGGCGCTGCACCGCGCGGTGCAAAATAGCCATGACCTTCAAAAGCTT 432
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTrpGluAspPhe 153
QY 433 TTTTACTGCTGAATCTTTGTAGAAACCATGAAAGACTTTCAAAGCGCTGGGAAGG 492
Db 154 LysTrpCysTrpGluAsnPheValTyTrpAsnAspAsnGluProPheLysProTrpLysGly 173
QY 493 CTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGGCATCCTT 540

```

Db 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189
RESULT 3
US-08-687-895-1
; Sequence 1, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
US-08-687-895-1
Alignment Scores:
Pred. No.: 1.58e-23 Length: 222
Score: 271.00 Matches: 63
Percent Similarity: 50.00% Conservative: 25
Best Local Similarity: 35.80% Mismatches: 76
Query Match: 24.61% Indels: 12
DB: 1 Gaps: 5
US-09-966-880a-7_COPY_80_676 (1-597) x US-08-687-895-1 (1-222)
QY 31 TTTCTTTACCAATGCTCCGCTAAGGCTAAGGCTGCGGTGAGACCTACCTCTGC 90
Db 52 PheLysPheGlnPheArgAsnValGluTyrSerSerGlyArgAsnLysThrPheLeuCys 71
QY 91 TAGGTAGTGAAGAGCGGTGACAGTGTACATCTTTTCTACTGGACTTGGTTATCTCTGC 150
Db 72 TyrValValGluAlaGlnGlyLysGlyGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91
QY 151 ATAACAAC---GGTGGCAGGTGAATGCTCTTCTCGGTACATCTCGGACTGGGAC 207
Db 92 AspGluHisAlaAlaAlaHisAlaGluAlaPhePheAsnThrIleLeuPro---Ala 110
QY 208 CTAGACCTGGCGGTGCTACCGCGTCACCTGGTTTACCTCTCGGAGCCCTCTAGCAC 267
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerProCysAlaAla 130
QY 268 TGTGCCGACATGTGCGCGGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTC 327

Db 131 CysAlaAspArgIle***LysThrLeuSerLysThrLysAsnLeuArgLeuLeuLeu 150
QY 328 ACCCGCGCTCTTACTTCTGTGTGAGACGCGCAAGGCTGAGCCGAG-----GGGCTG 378
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluIleGlnAlaAlaLeu 166
QY 379 CGCGCGCTGCACCGCGCGGTGCAATAGCATCATGACCTTCAAGATATTTTAC 438
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186
QY 439 TGCTGGAATACCTTTGTAGAAACCATGAA-----AGAACTTTCAAAGCCCTGGGAA 489
Db 187 ValTrpGlnAsnPheValGluGlnGluGlyGluSerLysAlaPheGlnProTrpGlu 206
QY 490 GGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATC 537
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIle 222
RESULT 4
US-09-040-482-1
; Sequence 1, Application US/09040482
; Patent No. 5916556
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,482
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/687,895
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
US-09-040-482-1
Alignment Scores:
Pred. No.: 1.58e-23 Length: 222
Score: 271.00 Matches: 63
Percent Similarity: 50.00% Conservative: 25
Best Local Similarity: 35.80% Mismatches: 76
Query Match: 24.61% Indels: 12
DB: 1 Gaps: 5

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0239 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 436941
US-08-816-241-3

Alignment Scores:
Pred. No.: 1.4e-20 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 22.25% Indels: 10
DB: 3 Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x US-08-816-241-3 (1-116)

QY 226 TACCGCGTCACCTGGTTACCTCGAGCCCTGCTAGCAC-----TGTCGCCGACAT 279
Db 7 TyrArgValThrPheIleSerTrpSerProcysPheSerTrpGlyCysAlaGlyGlu 26
QY 280 GTGGCCGACCTTCTCGAGGGAGAACCCCAACCTCAGTCAGGATCTTCACCGCGCCCTC 339
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
QY 340 TACTTCTGTGAGGACCGCAAGCTGAGCCC-----GAGGGCTGCGGGCGCTGCAC 390
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
QY 391 CGCGCGGGGTGCAATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACT 450
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheLeuTyrCysTrpAspThr 81
QY 451 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTT 510
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTyrAspGlyLeuGluHisSerGln 101
QY 511 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 7

US-09-040-482-3
Sequence 3, Application US/09040482
Patent No. 5916556
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Hawkins, Phillip R.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,482
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/687,895
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0109 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 436941
US-09-040-482-3

Alignment Scores:
Pred. No.: 1.4e-20 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 22.25% Indels: 10
DB: 3 Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-040-482-3 (1-116)

QY 226 TACCGCGTCACCTGGTTACCTCGAGCCCTGCTAGCAC-----TGTCGCCGACAT 279
Db 7 TyrArgValThrPheIleSerTrpSerProcysPheSerTrpGlyCysAlaGlyGlu 26
QY 280 GTGGCCGACCTTCTCGAGGGAGAACCCCAACCTCAGTCAGGATCTTCACCGCGCCCTC 339
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
QY 340 TACTTCTGTGAGGACCGCAAGCTGAGCCC-----GAGGGCTGCGGGCGCTGCAC 390
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
QY 391 CGCGCGGGGTGCAATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACT 450
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheLeuTyrCysTrpAspThr 81
QY 451 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTT 510
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTyrAspGlyLeuGluHisSerGln 101
QY 511 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 8

US-09-128-395-3
Sequence 3, Application US/09128395
Patent No. 6087108
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

```

; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 436941
US-09-128-395-3

```

```

Alignment Scores:
Pred. No.: 1,4e-20 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 22.25% Indels: 10
DB: 3 Gaps: 3

```

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-128-395-3 (1-116)

```

QY 226 TACGCGGTACCTGGTTACCTCCTCGAGCCCTGCTAGCAC-----TGTGCGCGACAT 279
Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26
QY 280 GTGCGCGACTTTCGCGAGGAACCCCAACCTCAGTCTGAGGATCTTCACGCGCGCTC 339
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
QY 340 TACTTCTGTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTCGCGCGCTGCAC 390
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
QY 391 CGCGCGGGTGCAGATAGCATCATGACCTTCAAGATATATTTTACTGCTGGAATACT 450
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81
QY 451 TTTGTAGAAAACCATGAAGAACTTCAAGACCTGGAAGGGTGCATGAAATTCAGTT 510
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGln 101
QY 511 CGTCTCTCCAGACGCTTCGGCGCATCCTT 540
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

```

RESULT 9

```

US-08-687-895-4
; Sequence 4, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.

```

```

; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1177798
US-08-687-895-4

Alignment Scores:
Pred. No.: 4,29e-17 Length: 236
Score: 216.50 Matches: 44
Percent Similarity: 59.52% Conservative: 31
Best Local Similarity: 34.92% Mismatches: 42
Query Match: 19.66% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-08-687-895-4 (1-236)

QY 103 AGCGGTACAGTGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGC----- 150
Db 33 ArgLysGluAlaCysLeuLeuTyrGluIleLysTrpGlyMetSerArgLysIleTrpArg 52
QY 151 -----AATAAGAACGGCTGC---CAGGTGGAATTGCTCTCTCTC---CGCTACATCTCG 198
Db 53 SerSerGlyLysAsnThrThrAsnHisValGluValAsnPheIleLysLysPheThrSer 72
QY 199 GACTGGGACCTAGACCTCGCCCTGCTACCGGCTCACCTGGTTCACCTCTCTGGAGCCCC 258
Db 73 GluArgAspPheHisProSerIleSerCysSerIleThrTrpPheLeuSerTrpSerPro 92
QY 259 TGCTACGACTGTCCCGACATGTGGCGGACTTTCTCGGAGGGGAACCCCAACCTCAGTCTG 318
Db 93 CysTrpGluCysSerGlnAlaIleArgGluPheLeuSerArgHisProGlyValThrLeu 112
QY 319 AGGATCTTCAACCCCGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGCTG 378
Db 113 ValIleTyrValAlaArgLeuPheTrpHisMetAspGlnGln---AsnArgGlnGlyLeu 131
QY 379 CGCGCGCTGCACCGCGCGGGTGCAGATAGCATCATGACCTTCAAGATATATTTTAC 438
Db 132 ArgAspLeuValAsnSerGlyValThrIleGlnIleMetArgAlaSerGluTyrTyrHis 151
QY 439 TGCTGGAATACTTTTGTGA 456

```

	Qy	103	AGGCGTGACAGRGCTACATCCCTTTTCATCGGACTTGTGGTTATCCTCGC-----	150
			::: ::: ::: :::	
D	b	33	ArcLysGLuAlacYcsLeuLeuTYrGlutlleLvstprGlyMetSerArqLylelrtpArg	52
			::: ::: ::: :::	

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:26:00 ; Search time 19.2225 Seconds
(without alignments)
6641.044 Million cell updates/sec

Title: US-09-966-880A-7_COPY_80_676

Perfect score: 1101
Sequence: 1 atggacagccttgatgaa.....ttcgtacttgggactttga 597

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09966880/runat_14062003_175526_10413/app_query.fasta.1.9493
-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09966880.ecgn.1.105 @runat_14062003_175526_10413
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLONG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	98.6	198	9	US-09-966-880A-8
2	1008	91.6	198	9	US-09-966-880A-2
3	390	35.4	384	10	US-09-729-674-174
4	369.5	33.6	222	10	US-09-925-300-1639

5	211	19.2	229	9	US-09-966-880A-36
6	102.5	9.3	18636	9	US-10-073-912-17
7	99.5	9.0	902	9	US-10-184-644-303
8	99.5	9.0	902	9	US-10-184-634-303
9	98.5	8.9	2558	9	US-10-184-644-103
10	98.5	8.9	2558	9	US-10-184-634-103
11	97.5	8.9	1572	9	US-10-184-644-65
12	97.5	8.9	1572	9	US-10-184-634-65
13	97	8.8	2103	9	US-10-184-644-319
14	97	8.8	2103	9	US-10-184-634-319
15	95.5	8.7	1819	9	US-10-184-644-39
16	95.5	8.7	1819	9	US-10-184-634-39
17	95	8.6	1788	9	US-10-184-644-405
18	95	8.6	1788	9	US-10-184-634-405
19	93.5	8.5	4640	9	US-10-184-644-75
20	93.5	8.5	4640	9	US-10-184-634-75
21	92.5	8.4	4040	9	US-10-123-155-425
22	92	8.4	2570	9	US-10-123-155-539
23	92	8.4	3150	9	US-10-184-644-81
24	92	8.4	3150	9	US-10-184-634-81
25	91.5	8.3	3877	9	US-10-184-644-263
26	91.5	8.3	3877	9	US-10-184-634-263
27	91	8.3	1422	9	US-10-123-155-117
28	91	8.3	1656	9	US-10-123-155-239
29	91	8.3	2163	9	US-10-123-155-127
30	90.5	8.2	3038	9	US-10-184-644-261
31	90.5	8.2	3038	9	US-10-184-634-261
32	90.5	8.2	3265	9	US-10-184-644-257
33	90.5	8.2	3265	9	US-10-184-634-257
34	90.5	8.2	3288	9	US-10-184-644-477
35	90.5	8.2	3288	9	US-10-184-634-477
36	90	8.7	591	9	US-10-060-036-165
37	90	8.2	2294	9	US-10-184-644-283
38	90	8.2	2294	9	US-10-184-634-283
39	90	8.2	3732	9	US-10-123-155-71
40	89.5	8.1	1395	9	US-10-123-155-3
41	89.5	8.1	2933	9	US-10-123-155-345
42	89	8.1	1257	9	US-10-184-644-365
43	89	8.1	1257	9	US-10-184-634-365
44	89	8.1	2362	9	US-10-184-644-15
45	89	8.1	2362	9	US-10-184-634-15

ALIGNMENTS

RESULT 1
US-09-966-880A-8
; Sequence 8, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966.880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-880A-8

Alignment Scores:

Pred. No.: 1.82e-102 Length: 198
Score: 1086.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.64% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-966-880A-8 (1-198)

```
QY 1 ATGACAGCCTCTTGATGAACCGAGGAAGTTTCTTTACCAATTCCAAATGTCGCTGG 60
Db 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
QY 61 GCTAAGGTCGGCTGAGACCTACCTGTCTACGTAGTGAAGAGCGGTGACAGTGTACA 120
Db 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
QY 121 TCTCTTTCACTGACTTTGGTTATCTTCGCAATAAGACGGTGCACGTGGAATGTCTC 180
Db 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
QY 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCTACCTGG 240
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 241 TTCACCTCTCGAGCGCTGCTGAGGATCTTACCGCGCCCTCTACTCTGTGAGGAGG 300
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCCCTCTACTCTGTGAGGAGCGCAAG 360
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 361 GCTGAGCCCGAGGGCTCGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACC 420
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
QY 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAACTTTCAAA 480
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
QY 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
QY 541 TTGCCCTCTGATGAGTTGATGACTTACGACAGCCATTTCTACTTTGGGACTT 594
Db 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
```

RESULT 2

US-09-966-880A-2
; Sequence 2, Application US/09966880A
; Patent No. US20020164743A1

GENERAL INFORMATION:

APPLICANT: Honjo, Tasuku
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966, 880A
CURRENT FILING DATE: 2001-09-28
PRIOR FILING DATE: 2000-03-28
PRIOR FILING DATE: 2000-03-28
PRIOR FILING DATE: 2000-03-28
PRIOR FILING DATE: 1999-12-27
PRIOR FILING DATE: 1999-06-24
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 198
TYPE: PRT

; ORGANISM: Mus musculus

US-09-966-880A-2

Alignment Scores:

Pred. No.: 1.7e-94 Length: 198
Score: 1008.00 Matches: 183
Percent Similarity: 95.94% Conservative: 6
Best Local Similarity: 92.89% Mismatches: 8
Query Match: 91.55% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-966-880A-2 (1-198)

```
QY 1 ATGACAGCCTCTTTCATGAACCGGAGGAGTTTCTTTACCAATTCCAAATGTCGCTGG 60
Db 1 MetAspSerLeuLeuMetLysGlnLysPheLeuTyrHisPheLysAsnValArgTrp 20
QY 61 GCTAAGGTCGGCGCTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
QY 121 TCTCTTTCACTGAGCTTTGGTTATCTTCGCAATAAGACGGTGCACGTGGAATGTCTC 180
Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeu 60
QY 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGCTCACCTGG 240
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 241 TTCACCTCTCGAGCGCTGCTGAGGATCTTACCGCGACATGTGGCGGACTTTCTCGGAGG 300
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGAGGAGCGCAAG 360
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 361 GCTGAGCCCGAGGGCTCGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACC 420
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140
QY 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAGAACTTTCAAA 480
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160
QY 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180
QY 541 TTGCCCTCTGATGAGTTGATGACTTACGACAGCGCATTTCTGTTCTGTTGGGA 591
Db 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgMetLeuGly 197
```

RESULT 3

US-09-729-674-174

; Sequence 174, Application US/09729674

; Patent No. US20010039335A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fecthel, Kim
APPLICANT: Genetics Institute, Inc.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

Patent NO. US20020164743A1
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382

;
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-966-880A-36

Alignment Scores:

Pred. No.:	5,07e-13	Length:	229
Score:	211.00	Matches:	49
Percent Similarity:	55.30%	Conservative:	24
Best Local Similarity:	37.12%	Mismatches:	49
Query Match:	19.16%	Indels:	10
DB:	9	Gaps:	4

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-966-880A-36 (1-229)

QY	70	CGCGGTGAGACCTACCTGTGCTACCTAGTGAAG-----AGCGGTGACAGTGCTACATCC	123
Db	33	ArgLysGluThrCysLeuLeuTyrgluileasntPrglyGlyArgHisSerVal-----	50
QY	124	TTTTTACGTGGACTTGGTTATCTTCGCAATAAGAACGCTGCCAGTGGAAATTCCTCTC	183
Db	51	-----TTPArgHisThrSerGlnAsnThrSerAsnHisValGluValAsnPhe	66
QY	184	CTC-----CGCTACATCTCGAGCTGGACCTAGACCTGCGCCCTGCTACCGCGTGCACCTGG	240
Db	67	LeuGluLysPheThrGluArgTyrPheArgProAsnThrArgCysSerIleThrTrp	86
QY	241	TTCACCTCCTCGAGCCCTCTAGACTGTGCGCGACATGCGCGACTTTCCTCGAGGG	300
Db	87	PheLeuSerTrpSerProCysGlyGluCysSerArgAlaIleThrGluPheLeuSerArg	106
QY	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360
Db	107	HisProTyrrValThrLeuPheIleTyrrIleAlaArgLeuTyrrHisThrAspGlnArg	126
QY	361	GCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGTGCAATAGCCATCAGCACC	420
Db	127	---AsnArgGlnGlyLeuArgAspLeuIleSerSerGlyValThrIleGlnIleMetThr	145
QY	421	TTCAAGATATTTTACTGCTGGAATACTTTTGTGA	456
Db	146	GluingluTyrcystyrCysTrpArgAsnPheVal	157

RESULT 6

US-10-073-912-17
; Sequence 17, Application US/10073912
; Publication No. US2003007703A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ34C1
; CURRENT APPLICATION NUMBER: US/10/073, 912
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 18636
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-073-912-17

Alignment Scores:

Pred. No.:	0.239	Length:	18636
Score:	102.50	Matches:	39

Percent Similarity:	35.62%	Conservative:	13
Best Local Similarity:	26.71%	Mismatches:	55
Query Match:	9.31%	Indels:	39
DB:	9	Gaps:	4

US-09-966-880A-7_COPY_80_676 (1-597) x US-10-073-912-17 (1-18636)

QY	20	ACCGGAGGAAGTTTCTTTTACCAATTCA---AAATGTCCGCTGGGCTAAGGTCGGCGTG	76
Db	14185	ThrGlyGlyAlaThrThrThrThrCysThrAlaAlaAlaGlyThrGlyAlaGlyThr	14204
QY	77	AGACCTACCTGTGCTACGTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTCACTGGACT	136
Db	14205	ThrGlyAlaCysAlaAla-----	14210
QY	137	TGTGTTATCTTTCGCAATAAGAACGCTGCCACGTGGAATTGCTTCTCCTCGCTACATCT	196
Db	14211	-----AlaCysAlaThrCysAlaCysThrGlyThrAlaAlaCys	14223
QY	197	CGGACTGGGACCTAGACCCCTGCGCTCTACCGGTCACCTGGTTACCTCCTCGA---	253
Db	14224	ThrThrGlyAlaThrThrGlyAlaCysAlaCysThrThrGlyThrCysAlaGlyCys	14243
QY	254	-----GCCCTGCTACGACTGTCCCGACATGTGG	283
Db	14244	CysAlaAlaThrGlyThrAlaGlyThrCysAlaCysAlaThrThrAlaAlaGlyAla	14263
QY	284	CCGACTTTCGCGAGGAAACCCCACTAGTCTGAGGATCTTCACCGCGCGCTCTACT	343
Db	14264	AlaThrAlaCysCysThrThrGlyThrAlaThrThrGlyAlaAlaAlaThrAlaThr	14283
QY	344	-----TCTGTGAGGACCGAAGGTGAGCCGAGGGCTGGCGGGC	385
Db	14284	GlyCysThrCysCysCysThrThrThrThrGlyThrThrCysThrGlyThrGlyGly	14303
QY	386	TGCACCGCGCGGGTGC	403
Db	14304	CysThrThrThrGlyCys	14309

RESULT 7

US-10-184-644-303
; Sequence 303, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 303
; LENGTH: 902
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-303

Alignment Scores:

Pred. No.:	0.192	Length:	902
Score:	99.50	Matches:	41
Percent Similarity:	37.69%	Conservative:	8
Best Local Similarity:	31.54%	Mismatches:	68

```
Query Match: 9.04% Indels: 13
DB: 9 Gaps: 5
US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-644-303 (1-902)
QY 23 GGAGAGAGTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCTGAGACCT 82
Db 432 GlyGlyGlyCysCysAlaGlyGlyCysAlaGlyThrGlyGlyGlyCysAlaThr 451
QY 83 ACCTGTGCTAGTAGAGGCGCTGACAGTGTCTACATCCTTTTCACTGGACTTTGGTT 142
Db 432 GlyGlyGlyCysCysAlaGlyGlyCysAlaGlyThrGlyGlyGlyCysAlaThr 451
QY 83 ACCTGTGCTAGTAGAGGCGCTGACAGTGTCTACATCCTTTTCACTGGACTTTGGTT 142
Db 452 ThrCysAlaThrGlyGlyAlaGlyAlaThrThrCys-----ThrCysCys 466
QY 143 ATCTTCGCAATAAAGACGGCTGCCACGTGGAATTCCTCTCCGCTACATCTCGGACT 202
Db 467 ThrCysAlaAlaThrThrCysThrThrCys---CysThrThrAlaThrThrCysAla 485
QY 203 GGGACTAGACCCCTGCGCTACCGCTGCTACCTGCTACCTGGTTCACCTCTCGAGCCCTGCT 262
Db 486 GlyCys---ThrThrThrCysAlaThrGlyAlaCysGlyCysThrGlyGlyThrCysAla 504
QY 263 ACGACTGTGCCGACATGTGCCGACTTCTGC-----GAGGGACCCCAACC 310
Db 505 ThrThrAla-----ThrCysThrThrGlyCysThrGlyCysAlaThrGlyThrAlaThr 522
QY 311 TCAGTCTGAGGATCTTCAACGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCG 370
Db 523 ThrCysThrGlyGlyGlyCysAlaThrThrGlyThrAlaThrThrThrThrThr 542
QY 371 AGGGCTCGCGCGCTGACCGCGCGGGG 400
Db 543 GlyAlaThrGlyGlyCysThrGlyThrGly 552
RESULT 8
US-10-184-634-303
; Sequence 303, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P343081C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 303
; LENGTH: 902
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-303
Alignment Scores:
Pred. No.: 0.192 Length: 902
Score: 99.50 Matches: 41
Percent Similarity: 37.69% Conservative: 8
Best Local Similarity: 31.54% Mismatches: 68
Query Match: 9.04% Indels: 13
DB: 9 Gaps: 5
US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-634-303 (1-902)
QY 23 GGAGAGAGTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCTGAGACCT 82
```

```
Db 432 GlyGlyGlyCysCysAlaGlyGlyCysAlaCysAlaGlyThrGlyGlyCysAlaThr 451
QY 83 ACCTGTGCTAGTAGAGGCGCTGACAGTGTCTACATCCTTTTCACTGGACTTTGGTT 142
Db 452 ThrCysAlaThrThrGlyGlyAlaGlyAlaThrThrCys-----ThrCysCys 466
QY 143 ATCTTCGCAATAAAGACGGCTGCCACGTGGAATTCCTCTCCGCTACATCTCGGACT 202
Db 467 ThrCysAlaAlaThrThrCysThrThrCys---CysThrThrAlaThrThrCysAla 485
QY 203 GGGACTAGACCCCTGCGCTACCGCTGCTACCTGCTACCTGGTTCACCTCTCGAGCCCTGCT 262
Db 486 GlyCys---ThrThrThrCysAlaThrGlyAlaCysGlyCysThrGlyGlyThrCysAla 504
QY 263 ACGACTGTGCCGACATGTGCCGACTTCTGC-----GAGGGACCCCAACC 310
Db 505 ThrThrAla-----ThrCysThrThrGlyCysThrGlyCysAlaThrGlyThrAlaThr 522
QY 311 TCAGTCTGAGGATCTTCAACGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCG 370
Db 523 ThrCysThrGlyGlyGlyCysAlaThrThrGlyThrAlaThrThrThrThrThr 542
QY 371 AGGGCTCGCGCGCTGACCGCGCGGGG 400
Db 543 GlyAlaThrGlyGlyCysThrGlyThrGly 552
RESULT 9
US-10-184-644-103
; Sequence 103, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P343081C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 103
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-103
Alignment Scores:
Pred. No.: 0.333 Length: 2558
Score: 98.50 Matches: 33
Percent Similarity: 32.81% Conservative: 9
Best Local Similarity: 25.78% Mismatches: 57
Query Match: 8.95% Indels: 29
DB: 9 Gaps: 3
US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-644-103 (1-2558)
QY 20 ACCGGAGAGTTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCTGAGA 79
Db 1734 ThrGlyGlyAlaAlaAlaAlaThrGlyAlaThrThrThrThrGlyAlaGly---GlyThrGly 1752
QY 80 CTTACCTGTGCTAGTAGTAGAGGCGGTGACAGTGTCTACATCCTTTTCACTGGACTTTG 139
Db 1753 ThrThrCysThrThr----- 1757
```

```
QY 140 GTTATCTTCGAATAAGAACGGCTGCCACGTGGAATTGCTCTTCTCCGCTACATCTCGG 199
Db 1758 -----CysCysAlaAlaCysGlyAlaCysThrThrGlyAlaAlaThr 1772
QY 200 ACTGGACCTAGACCTTGGCGGTGCTACCGGTCACCTGGTTCACCTCTGGAGGCCCT 259
Db 1773 ThrGlyCysThrThrCysAlaGlyGlyCysAlaGlyAlaGlyCysAlaCysGlyGlyThr 1792
QY 260 GCTACGACTGTGCCCGACATGTGCCGACATTTCTGGAGGGAACCCCAACCTCAGTCTGA 319
Db 1793 AlaThrAla-----CysThrAlaAlaAlaAlaAlaAlaAlaAlaThrThr 1804
QY 320 GGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGCTGC 379
Db 1805 GlyGlyGlyAlaAlaAlaCysAlaAlaAlaCysAlaAlaAlaAlaAlaAlaThrThr 1824
QY 380 GCGCGGTGCACCGCGCGGGTGC 403
Db 1825 GlyGlyCysThrAlaThrCysCys 1832
RESULT 10
US-10-184-634-103
; Sequence 103, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 103
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-103
Alignment Scores:
Pred. No.: 0.333 Length: 2558
Score: 98.50 Matches: 33
Percent Similarity: 32.81% Conservative: 9
Best Local Similarity: 25.78% Mismatches: 57
Query Match: 8.95% Indels: 29
DB: 9 Gaps: 3
US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-634-103 (1-2558)
QY 20 ACCGAGGAAGTTTCTTTACCAATTCAAAATGTCGCGTGAAGGTGCGGTGAGA 79
Db 1734 ThrGlyGlyAlaAlaAlaThrGlyAlaThrThrThrGlyAlaGly---GlyThrGly 1752
QY 80 CTTACTGTGCTACGTAGTAGTGAAGAGCGGTGACAGTGTCTACCTCTTTCACCTGGACTTG 139
Db 1753 ThrThrCysThrThr----- 1757
QY 140 GTTATCTTCGAATAGAGCGCTGCCACGTGGAATTGCTCTTCTCCGCTACATCTCGG 199
Db 1758 -----CysCysAlaAlaCysGlyAlaCysThrThrGlyAlaAlaThr 1772
QY 200 ACTGGACCTAGACCTTGGCGGTGCTACCGGTCACCTGGTTCACCTCTGGAGGCCCT 259
Db 1773 ThrGlyCysThrThrCysAlaGlyGlyCysAlaGlyAlaGlyCysAlaCysGlyGlyThr 1792
QY 260 GCTACGACTGTGCCCGACATGTGCCGACATTTCTGGAGGGAACCCCAACCTCAGTCTGA 319
Db 1793 AlaThrAla-----CysThrAlaAlaAlaAlaAlaAlaAlaAlaThrThr 1804
QY 320 GGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGCTGC 379
Db 1805 GlyGlyGlyAlaAlaAlaCysAlaAlaAlaCysAlaAlaAlaAlaAlaAlaThrThr 1824
QY 380 GCGCGGTGCACCGCGCGGGTGC 403
Db 1825 GlyGlyCysThrAlaThrCysCys 1832
RESULT 10
US-10-184-634-103
; Sequence 103, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 103
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-103
Alignment Scores:
Pred. No.: 0.333 Length: 2558
Score: 98.50 Matches: 33
Percent Similarity: 32.81% Conservative: 9
Best Local Similarity: 25.78% Mismatches: 57
Query Match: 8.95% Indels: 29
DB: 9 Gaps: 3
US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-634-103 (1-2558)
QY 20 ACCGAGGAAGTTTCTTTACCAATTCAAAATGTCGCGTGAAGGTGCGGTGAGA 79
Db 1734 ThrGlyGlyAlaAlaAlaThrGlyAlaThrThrThrGlyAlaGly---GlyThrGly 1752
QY 80 CTTACTGTGCTACGTAGTAGTGAAGAGCGGTGACAGTGTCTACCTCTTTCACCTGGACTTG 139
Db 1753 ThrThrCysThrThr----- 1757
QY 140 GTTATCTTCGAATAGAGCGCTGCCACGTGGAATTGCTCTTCTCCGCTACATCTCGG 199
Db 1758 -----CysCysAlaAlaCysGlyAlaCysThrThrGlyAlaAlaThr 1772
QY 200 ACTGGACCTAGACCTTGGCGGTGCTACCGGTCACCTGGTTCACCTCTGGAGGCCCT 259
Db 1773 ThrGlyCysThrThrCysAlaGlyGlyCysAlaGlyAlaGlyCysAlaCysGlyGlyThr 1792
QY 260 GCTACGACTGTGCCCGACATGTGCCGACATTTCTGGAGGGAACCCCAACCTCAGTCTGA 319
Db 1793 AlaThrAla-----CysThrAlaAlaAlaAlaAlaAlaAlaAlaThrThr 1804
QY 320 GGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGCTGC 379
Db 1805 GlyGlyGlyAlaAlaAlaCysAlaAlaAlaCysAlaAlaAlaAlaAlaAlaThrThr 1824
QY 380 GCGCGGTGCACCGCGCGGGTGC 403
Db 1825 GlyGlyCysThrAlaThrCysCys 1832
RESULT 11
US-10-184-644-65
; Sequence 65, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 65
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-65
Alignment Scores:
Pred. No.: 0.364 Length: 1572
Score: 97.50 Matches: 35
Percent Similarity: 33.33% Conservative: 10
Best Local Similarity: 25.93% Mismatches: 53
Query Match: 8.86% Indels: 37
DB: 9 Gaps: 4
US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-644-65 (1-1572)
QY 23 GGAGGAAGTTTCTTTACCAATTCAAAATGTCGCGTGAAGGTGCGGTGAGACCT 82
Db 357 GlyGlyAlaGlyGlyThrGlyAlaThrGlyCysAlaGlyThrAlaThrClyAlaCysThr 376
QY 83 ACCTGTGCTACGTAGTAGGAGCGGTGACAGTGTCTACATCTCTTTCACCTGGACTTTGGTT 142
Db 377 AlaCys----- 378
QY 143 ATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTCTTCTCCGCTACATCTCGGACT 202
Db 379 -----ThrAlaThrThrCysThrThrCysAlaThrAlaThrThrThrThr 393
QY 203 GGGACCTAGACCTGGCGCGTGTACCGCGTACCTGGTTCACCTCTGAGCCCTGCT 262
Db 394 GlyAla---ThrAlaThrAlaThrThrThrCysThrThrCysThrGlyCysAlaGly 412
QY 263 ACGACTGTGCCCGACATGTGCCGACATTTCTGGAGGGAACCCCAACCTCAGTCTGAGGA 322
Db 413 ThrThrThr-----ThrThrCysGlyAlaThrThrThrAlaAlaAlaGly 427
```



```
QY 323 TCTTACCAGCGGCTTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGTGC--- 379
    |||:|||||: ||| ||| ||| ||| |||
Db 428 ThrGlyThrAlaAlaThrAlaCysThrThrGlyCysAlaThrAlaThrGlyCysThr 447
    |||:|||||: ||| ||| ||| ||| |||
QY 380 -----GGCGGCTGACCGCGCGCGGGTGC 403
    ||| ||| ||| ||| |||
Db 448 GlyThrGlyThrGlyCysAlaGlyAlaCysThrGlyCysGlyCys 462

RESULT 12
US-10-184-634-65
; Sequence 65, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 65
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-65

Alignment Scores:
Pred. No.: 0.364 Length: 1572
Score: 97.50 Matches: 35
Percent Similarity: 33.33% Conservative: 10
Best Local Similarity: 25.93% Mismatches: 53
Query Match: 8.86% Indels: 37
DB: Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-634-65 (1-1572)
QY 23 GGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGGGCTAAGGCTGCGGTGAGACCT 82
    |||:|||||: ||| ||| ||| ||| |||
Db 357 GlyGlyAlaGlyThrGlyAlaThrGlyAlaGlyCysAlaGlyThrAlaThrGlyAlaCysThr 376
    |||:|||||: ||| ||| ||| ||| |||
QY 83 ACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTCACTGGACTTTGGTT 142
    ||| ||| ||| ||| ||| |||
Db 377 AlaCys-----ThrCys----- 378
QY 143 ATCTTCGCAATAAGACGGCTGCACGTGAATTGCTTCTTCCCTCCGCTACATCTCGGACT 202
    |||:|||||: ||| ||| ||| ||| |||
Db 379 -----ThrAlaThrThrCysThrThrCysAlaThrAlaThrThrThrThr 393
QY 203 GGGACCTAGACCTCGCCGCTGCTACCGGCTCACCTGGTTCACTCTCGGAGCCCTGCT 262
    ||| ||| ||| ||| ||| |||
Db 394 GlyAla---ThrAlaThrAlaThrThrThrCysThrThrCysThrGlyGlyCysAlaGly 412
QY 263 AGACTGTGCGGACATGCGGCGACTTTCTGCGAGGGAACCCCACTCAGTCTGAGGA 322
    ||| ||| ||| ||| ||| |||
Db 413 ThrThrThr-----ThrCysGlyAlaThrThrThrAlaAlaAlaGly 427
QY 323 TCTTACCAGCGGCTTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGTGC--- 379
    |||:|||||: ||| ||| ||| ||| |||
Db 428 ThrGlyThrAlaAlaThrAlaCysThrThrGlyCysAlaThrAlaThrGlyCysThr 447
    |||:|||||: ||| ||| ||| ||| |||
QY 380 -----GGCGGCTGACCGCGCGGGTGC 403
```

```
Db 448 GlyThrGlyThrGlyCysAlaGlyAlaCysThrGlyCysGlyCys 462
    ||| ||| ||| ||| ||| |||
RESULT 13
US-10-184-644-319
; Sequence 319, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 319
; LENGTH: 2103
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-319

Alignment Scores:
Pred. No.: 0.447 Length: 2103
Score: 97.00 Matches: 35
Percent Similarity: 36.13% Conservative: 8
Best Local Similarity: 29.41% Mismatches: 64
Query Match: 8.81% Indels: 12
DB: Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-644-319 (1-2103)
QY 56 GCTGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGGCGTGACAGTG 115
    |||:|||||: ||| ||| ||| ||| |||
Db 1568 AlaGlyCysCysThrGlyThrAlaThrThrCysAlaThr---ThrThrGlyThrThrCys 1586
    ||| ||| ||| ||| ||| |||
QY 116 CTACATCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGCGTGCCACGTGGGAAT 175
    ||| ||| ||| ||| ||| |||
Db 1587 -----ThrCysThrAlaGlyAlaAlaGlyThrThrThrGlyThr 1600
QY 176 TGCTCTTCCCTCCGCTACATCTCGGACTGGG-----ACCTAGACCTCGCCGCTGCT 226
    |||:|||||: ||| ||| ||| ||| |||
Db 1601 CysAlaGlyAlaAlaThrThrThrGlyAlaCysThrThrGlyThrGlyAlaCys 1620
    ||| ||| ||| ||| ||| |||
QY 227 ACCGCGTCACCTGCTTCACTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGGCG 286
    ||| ||| ||| ||| ||| |||
Db 1621 AlaThrAlaAlaAlaThrThrThrGlyThrAlaAlaThrGlyCysAlaThrAlaThrAla 1640
QY 287 ACTTTCGCGAGGGAACCCCACTCAGTCTGAGAGATCTTCACCGCGGCGCTCTACTTCT 346
    ||| ||| ||| ||| ||| |||
Db 1641 ThrAlaCysAlaAlaThrThrThrGlyAlaAlaGlyCysAlaCysThrCysCysThrThr 1660
    ||| ||| ||| ||| ||| |||
QY 347 GTGAGGACCGCAAGGCTGAGCCCGGCGGCTGCGGCGCTGCACCGCGCGGGTGC 403
    ||| ||| ||| ||| ||| |||
Db 1661 ThrThrCysThrThrCysAlaGlyThrThrCys-----CysThrCysAlaGlyCys 1677

RESULT 14
US-10-184-634-319
; Sequence 319, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
```

```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 319
; LENGTH: 2103
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-319

Alignment Scores:
Pred. No.: 0.447          Length: 2103
Score: 97.00             Matches: 35
Percent Similarity: 36.13% Conservative: 8
Best Local Similarity: 29.41% Mismatches: 64
Query Match: 8.81%       Indels: 12
DB: 9                     Gaps: 4

US-09-966-880a-7_COPY_80_676 (1-597) x US-10-184-634-319 (1-2103)
QY 56 GCTGGCTAAGGTCGGCTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTG 115
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1568 AlaGlyCysCysThrGlyThrAlaThrThrCysAlaThr---ThrThrGlyThrThrCys 1586
QY 116 CTACATCCTTTTCACTGGACTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAA 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1587 -----ThrCysThrAlaGlyAlaAlaGlyThrThrThrGlyThr 1600
QY 176 TGCTCTTCTCGGTACATCTCGGACTGG-----ACCTAGACCTGGCGGCTGCT 226
    ||||| :|||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 1601 CysAlaGlyAlaAlaThrThrThrThrThrGlyAlaCysThrThrGlyThrThrGlyAlaCys 1620
QY 227 ACCGGGTACCTGGTTCACCTCCCTGGAGCCCTGCTAGACTGTGCCGACATGTGGCG 286
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1621 AlaThrAlaAlaAlaThrThrThrThrAlaAlaThrGlyCysAlaThrAlaThrAla 1640
QY 287 ACTTCTCGAGGGAACCCCAACCTCAGTCGAGGATCTTCACCGCGGCTCTACTTCT 346
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1641 ThrAlaCysAlaAlaThrThrThrGlyAlaAlaGlyCysAlaCysThrCysThrThr 1660
QY 347 GTGAGGACCGCAGGCTGAGCCCGAGGGGCTGGCGGCTGCACCGCGCGGGGTGC 403
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1661 ThrThrCysThrThrCysAlaGlyThrThrCys-----CysThrCysAlaGlyCys 1677

RESULT 15
US-10-184-644-39
; Sequence 39, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

```
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 39
; LENGTH: 1819
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-39

Alignment Scores:
Pred. No.: 0.609          Length: 1819
Score: 95.50             Matches: 41
Percent Similarity: 38.97% Conservative: 12
Best Local Similarity: 30.15% Mismatches: 62
Query Match: 8.67%       Indels: 21
DB: 9                     Gaps: 6
```

```
US-09-966-880a-7_COPY_80_676 (1-597) x US-10-184-644-39 (1-1819)
QY 20 ACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGGTGAGA 79
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1477 ThrGlyGlyAlaAlaCysGlyAlaGlyThrThrThrGly-----AlaGlyThrAla 1494
QY 80 CCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCACTGGACTTTG 139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1495 AlaThrCysAlaGlyGlyAlaAlaGlyThrAla-----ThrAla 1507
QY 140 GTTATCTTCGAATAAGAACGGCTGCCACGTGGGAATGCTCTTCTCGGTACATCTCGG 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1508 ThrCysThrAlaThrAlaThrGlyAlaThr-----CysThrThrGlyAlaThrAlaThr 1525
QY 200 ACTGGACCTAGACCCCTGGCC---GCTGCTACCGCTCACCTGGTTCACCTCCTGGAGCC 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1526 ThrGlyThrThrThrAlaThrAlaThrAlaAlaThrThrThrGlyAlaAlaGly 1545
QY 257 CCTGCTACGACTGTGCCCGGACATGTGCCCGACTTTCTGCGAGGGAACCCCAAGCTCAGTC 316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1546 ThrCysThrAlaAlaAlaAlaGlyAlaCysThrGlyCysAlaThrThrThrThr--- 1564
QY 317 TGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGC 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 ---AlaAlaAlaCysAlaAlaGlyThrThrAlaGlyThrAlaThrThrAlaThrGly 1583
QY 377 TGGCGC-----GCTGCACCGCGCGGGTGC 403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1584 CysGlyThrThrGlyCysCysAlaCysGlyThrAlaGlyCys 1599
```

Search completed: June 14, 2003, 19:09:17
Job time : 38.2225 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 17:59:23 ; Search time 71 Seconds
(without alignments)
371.600 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLMMNRKFLYQFNVRW.....ILLPLYEVDLDRFRTGL 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1086	100.0	198	21	Human activation-i
2	1008	92.8	198	21	Mouse activation-i
3	390	35.9	384	20	Amino acid sequenc
4	390	35.9	384	22	Human secreted pro
5	390	35.9	384	23	Human polypeptide
6	388	35.7	384	21	Amino acid sequenc
7	369.5	34.0	190	19	Human RNA editing
8	369.5	34.0	190	21	Human RNA editing
9	369.5	34.0	190	22	Human polypeptide
10	369.5	34.0	210	22	Novel human enzyme

11	369.5	34.0	221	22	Human polypeptide
12	369.5	34.0	222	21	Human prostate can
13	369.5	34.0	222	22	Novel human enzyme
14	336.5	31.0	268	23	Human RNA metaboli
15	321.5	29.6	272	22	Novel human enzyme
16	309.5	28.5	163	23	Human secretory po
17	274	25.2	224	23	Human AAD45360 pro
18	274	25.2	224	23	Human APOBEC2 prot
19	271	25.0	222	19	A novel human mRNA
20	262.5	24.2	367	22	Human polypeptide
21	208.5	19.2	261	22	Novel human secret
22	204.5	18.8	169	22	Novel human secret
23	203	18.7	113	21	Human secreted pro
24	199	18.3	236	15	Apo-B RNA editing
25	198	18.2	127	22	Human protein sequ
26	193	17.8	229	15	Apo-B RNA editing
27	103.5	9.5	94	21	Human secreted pro
28	102	9.4	720	22	Drosophila melanog
29	84	7.7	51	22	Peptide #5883 enco
30	84	7.7	51	22	Protein #5554 enco
31	84	7.7	51	22	Human brain expres
32	84	7.7	51	22	Human bone marrow
33	84	7.7	51	22	Peptide #5602 enco
34	84	7.7	51	22	Peptide #5853 enco
35	84	7.7	51	23	Human peptide enco
36	81.5	7.5	2342	22	Drosophila melanog
37	80.5	7.4	410	22	Human transferrase
38	78.5	7.2	402	22	Pseudorabies virus
39	77.5	7.1	261	23	Human caspase-12 i
40	77.5	7.1	402	15	Glycoprotein 50 (g
41	77.5	7.1	402	21	Novel recombinant
42	77.5	7.1	402	22	Pseudorabies virus
43	77.5	7.1	402	22	Pseudorabies virus
44	77.5	7.1	402	22	Pseudorabies virus
45	77.5	7.1	663	22	Human PRO-C-MG.72

ALIGNMENTS

RESULT 1

AAAB24198

ID AAAB24198 standard; Protein: 198 AA.

AC AAAB24198;

DT 05-FEB-2001 (first entry)

DE Human activation-induced cytidine deaminase SEQ ID NO:8.

Activation-induced cytidine deaminase; AID; cytidine deaminase;
immune related disease; allergy; allergic disease; antiallergic;
antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;
gene therapy; B cell associated immune system disorder; food allergy;
immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
ataxia telangiectasia; common variable immunodeficiency disorder;
major histocompatibility class II deficiency disease;
auto immunodeficiency syndrome; IgG subclass selection disorder.

OS Homo sapiens.

PN WO2000058480-A1.

PD 05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP01918.

PR 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

PR 27-DEC-1999; 99JP-0371382.

XX

```
PA (NISB ) JAPAN TOBACCO INC.
XX (HONJ/) HONJO T.
PI Honjo T, Muramatsu M;
XX WPI; 2000-611715/58.
XX N-PSDB; AAC55312.
XX Nucleic acid encoding activation induced cytidine deaminase, useful as
XX a target for drug development for immune-related diseases including
XX allergies -
XX Claim 1; Page 140-141; 174pp; Japanese.
XX The present sequence is human activation-induced cytidine deaminase
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
XX has cytidine activity similar to APOBEC-1. AID has anti-allergic,
XX antianemic, antiasthmatic, ophthalmological, anti-HIV and
XX dermatological activities, and can be used in gene therapy. AID
XX polynucleotides are useful in methods for identifying drugs for the
XX treatment of B cell associated immune system disorders, immunodeficiency
XX diseases and allergies, such as immunoglobulin A (IgA) deficiency
XX colitis, asthma, food allergy, drug allergy, allergic rhinitis, allergic
XX disease, DiGeorge disease, ataxia telangiectasia, common variable
XX immunodeficiency disorder, MHC (major histocompatibility class) class
XX II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
XX IgE disorder, and IgG subclass selection disorder. The DNA sequences
XX encoding AID may be used for gene therapy and the antibodies to the AID
XX protein may be used for diagnosis and treatment of these disorders.
XX Sequence 198 AA;
SQ Query Match 100.0%; Score 1086; DB 21; Length 198;
Best Local Similarity 100.0%; Pred. No. 7.1e-116;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
DB 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
QY 61 FLRYISDWLDPGRCYRVWTFTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120
DB 61 FLRYISDWLDPGRCYRVWTFTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120
QY 121 AEPEGLRLRHAGVOIAIMTFKDYFCYCNWTFVENHRTFKAWEGHLSVRLSRLRL 180
DB 121 AEPEGLRLRHAGVOIAIMTFKDYFCYCNWTFVENHRTFKAWEGHLSVRLSRLRL 180
QY 181 LPLYEVDLDRDAFRTLGL 198
DB 181 LPLYEVDLDRDAFRTLGL 198
RESULT 2
AAB24197
ID AAB24197 standard; Protein; 198 AA.
XX AAB24197;
AC AAB24197;
XX 05-FEB-2001 (first entry)
XX Mouse activation-induced cytidine deaminase SEQ ID NO:2.
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;
XX immune related disease; allergy; allergic disease; anti-allergic;
XX antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatologic;
XX gene therapy; B cell associated immune system disorder; food allergy;
XX immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
XX IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
XX drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
XX ataxia telangiectasia; common variable immunodeficiency disorder;
XX major histocompatibility class II deficiency disease;
```

```
KW auto immunodeficiency syndrome; IgG subclass selection disorder.
XX Mus musculus.
XX WO200058480-A1.
XX 05-OCT-2000.
XX 28-MAR-2000; 2000WO-JP01918.
XX 29-MAR-1999; 99JP-0087192.
XX 24-JUN-1999; 99JP-0178999.
XX 27-DEC-1999; 99JP-0371382.
XX (NISB ) JAPAN TOBACCO INC.
XX (HONJ/) HONJO T.
XX Honjo T, Muramatsu M;
XX WPI; 2000-611715/58.
XX N-PSDB; AAC55307.
XX Nucleic acid encoding activation induced cytidine deaminase, useful as
XX a target for drug development for immune-related diseases including
XX allergies -
XX Claim 1; Page 131-132; 174pp; Japanese.
XX The present sequence is mouse activation-induced cytidine deaminase
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
XX has cytidine activity similar to APOBEC-1. AID has anti-allergic,
XX antianemic, antiasthmatic, ophthalmological, anti-HIV and
XX dermatological activities, and can be used in gene therapy. AID
XX polynucleotides are useful in methods for identifying drugs for the
XX treatment of B cell associated immune system disorders, immunodeficiency
XX diseases and allergies, such as immunoglobulin A (IgA) deficiency
XX colitis, asthma, food allergy, drug allergy, allergic rhinitis, allergic
XX disease, DiGeorge disease, ataxia telangiectasia, common variable
XX immunodeficiency disease, MHC (major histocompatibility class) class
XX II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
XX IgE disorder, and IgG subclass selection disorder. The DNA sequences
XX encoding AID may be used for gene therapy and the antibodies to the AID
XX protein may be used for diagnosis and treatment of these disorders.
XX Sequence 198 AA;
SQ Query Match 92.8%; Score 1008; DB 21; Length 198;
Best Local Similarity 92.9%; Pred. No. 6e-107;
Matches 183; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
QY 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
DB 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
QY 61 FLRYISDWLDPGRCYRVWTFTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120
DB 61 FLRYISDWLDPGRCYRVWTFTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120
QY 121 AEPEGLRLRHAGVOIAIMTFKDYFCYCNWTFVENHRTFKAWEGHLSVRLSRLRL 180
DB 121 AEPEGLRLRHAGVOIAIMTFKDYFCYCNWTFVENHRTFKAWEGHLSVRLSRLRL 180
QY 181 LPLYEVDLDRDAFRTLGL 197
DB 181 LPLYEVDLDRDAFRTLGL 197
RESULT 3
AAY42383
ID AAY42383 standard; Protein; 384 AA.
XX AAY42383;
AC AAY42383;
```

Qy	116	CEDRKAPEGLRLRHACGVAITMTFKDYFCWNTFVENHRTFFKAWBGLHENSVRLSQ	179
Dd	316	-DDOGRQCQEGUPLTAEAGAKISIMTYSEFKHCWDTFVDHQGCFPFOWDGLDEHSODLSGR	374
Qy	176	LRRIL 180	
Dd	375	LRAIL 379	
RESULT 4			
AAU39075	ID	AAU39075 standard; Protein: 384 AA.	
XX	AC	AAU39075;	
XX	DT	16-JAN-2002 (first entry)	
XX	DE	Human secreted protein lp547_4.	
XX	KW	Human; secreted protein; antiinflammatory; immunosuppressive;	
XX	KW	nootropic; neuroprotective; antiarthritic; antimicrobial; vulnery;	
XX	KW	cytostatic; antidiabetic; virucide; antinfertility; anticonvulsant;	
XX	KW	vasotropic; antiparkinsonian; immunostimulant; dermatological;	
XX	KW	antirheumatic; anticumor; antitumor; osteopathic; tranquiliser;	
XX	KW	cerebroprotective; cytokine; cell proliferation; cell differentiation;	
XX	KW	immune deficiency; severe combined immunodeficiency; SCID; tumour;	
XX	KW	autoimmune disorder; multiple sclerosis; rheumatoid arthritis;	
XX	KW	graft-versus-host disease; myeloid deficiency; wound healing; ulcer;	
XX	KW	periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;	
XX	KW	Parkinson's disease; Huntington's disease; infection; cardiac disease;	
XX	KW	stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;	
XX	KW	food supplement; vaccine.	
XX	OS	Homo sapiens.	
XX	PN	WO200175068-A2.	
XX	PD	11-OCT-2001.	
XX	PP	22-MAR-2001; 2001WO-US09369.	
XX	PR	30-MAR-2000; 2000US-0539330.	
XX	PR	04-DEC-2000; 2000US-0729567A.	
XX	PA	(GBMY) GENETICS INST INC.	
XX	PI	Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;	
XX	PI	Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;	
XX	PI	Clark H, Fechtel K, Merberg D;	
XX	DR	WPI: 2001-6393363/73.	
XX	DR	N-PSDB; AAS59293.	
XX	PT	Secreted human proteins, useful as vaccine for treating various	
XX	PT	diseases such as autoimmune disorders (e.g. multiple sclerosis), and	
XX	PT	nervous system disorders (e.g. stroke) -	
XX	PS	Disclosure; Page 580-581; 619pp; English.	
XX	CC	The invention relates to novel human secreted proteins, the nucleic	
XX	CC	acids encoding them. The protein may exhibit cytokine, cell proliferation	
XX	CC	or cell differentiation activity or may induce production of other	
XX	CC	cytokines in certain cell populations and may exhibit immune stimulating	
XX	CC	or immune suppressing activity, which is useful for the treatment of	
XX	CC	various immune deficiencies and disorders e.g. severe combined	
XX	CC	immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,	
XX	CC	systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary	
XX	CC	inflammation. The proteins are also useful in the treatment of diseases	
XX	CC	and disorders including tissue, skin and organ transplantation and in	
XX	CC	graft-versus-host diseases (GVHD), in the induction of tumour immunity,	
XX	CC	myeloid or lymphoid cell deficiencies, wound healing and tissue repair,	
XX	CC	in the treatment of burns, incisions and ulcers; as well as in treatment	
XX	CC	of periodontal disease, osteoporosis or osteoarthritis, mediated by	

US2001039335-A1.

08-NOV-2001.

04-DEC-2000; 2000US-0729674.

10-APR-1997; 97US-136425P.

04-DEC-1997; 97US-067454P.

20-DEC-1997; 97US-068379P.

02-JAN-1998; 98US-070346P.

07-JAN-1998; 98US-070643P.

08-JAN-1998; 98US-070755P.

13-JAN-1998; 98US-071304P.

22-JAN-1998; 98US-072134P.

30-JAN-1998; 98US-073095P.

18-FEB-1998; 98US-075038P.

30-MAR-2000; 2000US-0539330.

The invention relates to isolated polynucleotides (ABA90876-ABA90968 and ABA90980) and encoded proteins (ABB55698-ABB55800), especially polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90985) and proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in clones bd306-7 and y8-1 respectively and the clones bd306-7 and y8-1 are deposited with the American Type Culture Collection (ATCC) with accession number 98599. The polynucleotides and encoded polypeptides have cytostatic, anti-inflammatory, immunomodulator, vulnery, neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic and anti-inflammatory activity and acting as cytokine modulators, haematopoiesis regulators, tissue growth modulators and/or cadherin suppressors. The polypeptides and polynucleotides are useful in gene therapies, particularly for preventing, treating or ameliorating any of the following diseases: immune deficiency and disorders; e.g. bacterial or fungal infections, autoimmune disorders, cancer, systemic lupus erythematosus or graft-versus-host disease; myeloid or lymphoid cell deficiencies; wound, burns, incisions and ulcers, osteoporosis or osteoarthritis; central and peripheral nervous system diseases and neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis or Shy-Drager syndrome; haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis or systemic inflammatory response syndrome, ischaemia-reperfusion injury, endotoxin lethality, arthritis, inflammatory bowel disease or Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus foliaceus.

SQ	Sequence	384 AA:
	Query Match	35.9%; Score 390; DB 23; Length 384;
	Best Local Similarity	44.9%; Pred. No. 7 5e-36;
	Matches	83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;
QY	6	MNRKFLYQPKNFWAKGRRETYLYVVKRRDSATSFSLD--FGYLRNK-----NGC 55
DB	197	MDPTFTFNENPWRGHRHETLYCEVERMHNDTWLLNQRRGLFCNOAPHKHGFLEGR 256
QY	56	HVELLFRLYISDDWLDOPGRCRYRVTFWTSNSPCYDCARHWADFLRGNPNLSRIETARLYF 115
DB	257	HAELCFDLVIPFNKKLDDQDYRVCTFSNSPCFCAQEMAKFI SKNNKVUSLCIFARIY- 315
QY	116	CEDRKAEPEGLRLRLRHAGVOIAIMTKDFYCWNTFVENHERTFKAMEGLLHENSVRLSRQ 175
DB	316	-DOGRCCOBLRLTLAAGAKISITSEFKHCWDTFVDHOGCPFPQMDGPDMDHSDQLSGR 374

[illegible]

PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-024617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 03-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-465566/50.
 N-PSDB; AAS41420.

XX Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 11; SEQ ID NO 1546; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 210 AA;
 Query Match 34.0%; Score 369.5; DB 22; Length 210;
 Best Local Similarity 44.9%; Pred. No. 7.3e-34;
 Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;
 QY 11 FLYQFKVNRWAKGRRETYLCYVK--RRDSATSFSLDFGYLRN---KNGCHVELLFLRY 64
 Db 37 FYQFKNLWEANDRNETWLCFTVEGIKRRSVWSKT--GVFRNQVDSETHCAERCFLSW 94
 QY 65 ISDWLDPGRCYRTWFTSPCYDCARHVADFLGNPNLSLRIFRTARLYFCEDRAEPE 124
 Db 95 FCDILSPNTKYQVTWYTSWSPDCAGEAEFLARHSNVLTIPTARLYFQ-YPCYQE 153
 QY 125 GLRLHRAGVOIAIMTFKDYFCWNTFVENHRTFKAWEGHLSNLSRLRRIL 180
 Db 154 GLRSLQEGVAVEIMDYEDFKYCWENFYNDNEPFPKWLKTNFRLLRRLRESL 209
 RESULT 11
 ID AAM40691 standard; Protein; 221 AA.
 XX
 AC AAM40691;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5622.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AAI59847.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 2; SEQ ID NO 5622; 10078pp; English.
 PS
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX Sequence 221 AA;

Query Match 34.0%; Score 369.5; DB 22; Length 221;
 Best Local Similarity 44.9%; Pred. No. 7.8e-34;
 Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;
 QY 11 FLYQFNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN---KNGCHVELLFLRY 64
 Db 48 FYFOFKNLWEANDRNELTLCFTVEGIKRRSVVSNKT--GVFRNQVDSETHCAERCFLSW 105
 QY 65 ISDWDLDPGRCYRVWFTSWSPCYDCARHVADELGRNPNLSLRIFTARLYFCEDRKAEP 124
 Db 106 FCDLILSPNTKYQVTWYTSWSPCDGAGEVAEFLARHSNVNLTIFTARLYFQ-YPCYQE 164
 QY 125 GLRLRHAGVQIAIMTFKDYFYCWNFTFVENHRTFFKAWEGHLSNVLSQLRRIL 180
 Db 165 GLRSLSQGVAVEIMDYEDFKYCWENFVNDNEFPFKWGLKTNFRLLKRLRESL 220

RESULT 12
 AAB57061
 ID AAB57061 standard; Protein: 222 AA.
 XX
 AC AAB57061;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human prostate cancer antigen protein sequence SEQ ID NO:1639.
 XX
 KW Human; prostate cancer; antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;

KW wound; infectious disease.
 XX
 OS Homo sapiens.
 XX
 PN WO2000055174-A1.
 XX
 PD 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US05988.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2000-587513/55.
 DR N-PSDB; AAF16264.
 XX
 XX Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 11; Page 2097-2098; 2338pp; English.
 XX
 CC AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.

Query Match 34.0%; Score 369.5; DB 21; Length 222;
 Best Local Similarity 44.9%; Pred. No. 7.9e-34;
 Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;
 QY 11 FLYQFNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN---KNGCHVELLFLRY 64
 Db 49 FYFOFKNLWEANDRNELTLCFTVEGIKRRSVVSNKT--GVFRNQVDSETHCAERCFLSW 106
 QY 65 ISDWDLDPGRCYRVWFTSWSPCYDCARHVADELGRNPNLSLRIFTARLYFCEDRKAEP 124
 Db 107 FCDLILSPNTKYQVTWYTSWSPCDGAGEVAEFLARHSNVNLTIFTARLYFQ-YPCYQE 165
 QY 125 GLRLRHAGVQIAIMTFKDYFYCWNFTFVENHRTFFKAWEGHLSNVLSQLRRIL 180
 Db 166 GLRSLSQGVAVEIMDYEDFKYCWENFVNDNEFPFKWGLKTNFRLLKRLRESL 221

RESULT 13
 AAU23537
 ID AAU23537 standard; Protein: 222 AA.
 XX
 AC AAU23537;
 XX
 DT 17-DEC-2001 (first entry)
 XX
 DE Novel human enzyme polypeptide #623.
 XX
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;

KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX Homo sapiens.
PN WO200155301-A2.
XX 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179055.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465566/50.
XX N-PSDB; AAS41407.
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
XX Claim 11; SEQ ID No 1533; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 222 AA;
SQ
Query Match 34.0%; Score 369.5; DB 22; Length 222;
Best Local Similarity 44.9%; Pred. No. 7.9e-34;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;
QY 11 FLYQFNKRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
Db 49 FYFQFNLEANDRNTELCTVEGIKRSVWSMKT--GVFRNQVDSETHCHAEKFLSW 106
QY 65 ISDWLDPGRCYRVWFTSWSPCYDCARHVADEPLRGNPNLSLRIFTARLYFCEDRKAEP 124
Db 107 FCDLILSPNTKYQVTWYTSWSPCCGAGEVAEFLARHSNVNLTIFTARLYFQ-YPCYQE 165
QY 125 GLRRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGLHENSRLSRLRRL 180
Db 166 GLRSLSQEGAVAEIMDYEDFKYCWENFVYNDNEPFPKWLKTNFRLKRLRESL 221
RESULT 14
AAE15256
ID AAE15256 standard; Protein; 268 AA.
XX
AC AAE15256;
XX
XX 07-MAR-2002 (first entry)
XX Human RNA metabolism protein-19 (RMEP-19).
XX
XX Human; RNA metabolism protein-19; RMEP-19; gout; nervous system disorder;
KW autoimmune; inflammatory; cell proliferative; developmental; thyroiditis;
KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;
KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;
KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;
KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;
KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;

KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; antitumor;
KW tranquiliser; drug screening; pancreatitis; renal tubular acidosis;
KW systemic lupus erythematosus; colitis; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Peptide 1..21
FT /label= Signal_peptide
FT Protein 22..268
FT /label= Mature_RMEP_19_protein
XX
XX WO200183524-A2.
XX
XX 08-NOV-2001.
XX
XX 27-APR-2001; 2001WO-US13862.
XX
XX 28-APR-2000; 2000US-200184P.
PR 04-MAY-2000; 2000US-201875P.
PR 04-MAY-2000; 2000US-202090P.
PR 06-JUN-2000; 2000US-210232P.
PR 25-JUL-2000; 2000US-220553P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;
PI Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;
XX
XX WPI; 2002-034502/04.
DR N-PSDB; AAD24392.
XX
XX New human RNA metabolism protein for diagnosing or treating nervous
PT system disorders, autoimmune/inflammatory disorders, cell proliferative
PT disorders and developmental disorders
XX
XX Claim 1; Page 145-146; 196pp; English.
XX
XX The invention relates to human RNA metabolism proteins (RMEP) and their
CC corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,
CC treating and preventing nervous system disorders (epilepsy, dementia,
CC stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);
CC prion diseases; fatal familial insomnia, nutritional and metabolic
CC diseases of the nervous system; inherited, metabolic, endocrine and
CC toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)
CC amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-
CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,
CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,
CC ulcerative colitis, and infections); cell proliferative disorders (cancer
CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental
CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening
CC techniques, gene therapy and for creating transgenic animals. The present
XX sequence is human RMEP-19 protein.

SQ Sequence 268 AA;

Query Match 31.0%; Score 336.5; DB 23; Length 268;
Best Local Similarity 40.8%; Pred. No. 6.1e-30;
Matches 71; Conservative 27; Mismatches 30; Indels 5; Gaps 2;

QY 11 FLYQFNKRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRN----KNGCHVELLFLRYIS 66
Db 95 FYFQFNLEANDRNTELCTVEGIKRSVWSMKT--GVFRNQVDSETHCHAEKFLSW 154
QY 67 DWDLDPGRCYRVWFTSWSPCYDCARHVADEPLRGNPNLSLRIFTARLYFCEDRKAEPGL 126
Db 155 DDILSPNTKYQVTWYTSWSPCCGAGEVAEFLARHSNVNLTIFTARLYFQYDQY-BGL 213
QY 127 RRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGLHENSRLSRLRRL 180
Db 214 CSLSQEGASVKIMGYKDFVSCWKNFYSDDEPFPKWLKTNFRLKRLRESL 267

RESULT 15
AAU23799
ID AAU23799 standard; Protein; 272 AA.
XX AC AAU23799;
XX DT 18-DEC-2001 (first entry)
XX DE
XX DE Novel human enzyme polypeptide #885.
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX OS Homo sapiens.
XX PN WO200153301-A2.
XX PD
XX PF 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01239.
XX PF 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226686.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 05-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-465566/50.
XX N-PSDB; AAS41669.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 11; SEQ ID No 1795; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 272 AA;
Query Match 29.6%; Score 321.5; DB 22; Length 272;
Best Local Similarity 45.3%; Pred. No. 3.2e-28;
Matches 68; Conservative 22; Mismatches 51; Indels 9; Gaps 4;
QY 11 FLYQFNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGLRN----KNGCHVELFLRY 64
DB 49 FYFQFNLRWANDRNETWLCFTVEGIKRSVYSWKT--GVFRNQVDSETHCAERCFLSW 106
QY 65 ISDWLDPCRCYRVTFWFTSWSPCYDCARHVADFLGNPNLSLRIFTARLYFCEDRKAEPE 124
DB 107 FCDLILSPNTKQVNTWYTSWSPCPDAGEVAEFLARHSNVLNLTIFTARLYFYQ-YPCYOE 165
QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVEN 154
DB 166 GLRSLSQEGVAEIMDYEDFKYCWENFVN 195

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:07:08 ; Search time 41 seconds
(without alignments)
464,259 Million cell updates/sec

Title: US-09-966-880A-8
Perfect score: 1086
Sequence: 1 MDSLMLNRRKFLYQFNVRW.....ILLPLYEVDLDRFAFTLGL 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	245	22.6	116	2 G01233	phorbolins I - huma
2	225	20.7	236	2 A53853	apolipoprotein B m
3	216.5	19.9	236	2 I59323	apolipoprotein B m
4	211	19.4	229	2 JC4269	apolipoprotein B m
5	211	19.4	229	2 I48249	apolipoprotein B m
6	198	18.2	229	2 I59577	apolipoprotein B m
7	97.5	9.0	369	2 S63464	hypothetical prote
8	81.5	7.5	2342	2 T13412	hypothetical prote
9	80	7.4	476	2 C64601	hypothetical prote
10	78.5	7.2	817	2 S53921	hypothetical prote
11	77.5	7.1	251	2 T32318	hypothetical prote
12	77.5	7.1	328	2 T30999	hypothetical prote
13	77.5	7.1	402	1 VGBE50	glycoprotein D pre
14	77.5	7.1	618	2 G59432	MacGAP protein [im
15	77	7.1	1829	2 T14280	RW1 protein - mous
16	76.5	7.0	466	1 S15280	phosphogluconate d
17	75.5	7.0	382	2 A47477	hyaluronidase - ho
18	75.5	7.0	533	2 A47143	retinal pigment mi
19	75	6.9	320	2 E70355	conserved hypothet
20	75	6.9	617	2 S52797	mufl protein - hum
21	74.5	6.9	229	2 F97193	single-strand DNA-
22	74	6.8	319	2 AC3455	UDP-galactose-lipi
23	74	6.8	307	2 S78538	site-specific reco
24	73.5	6.8	337	2 A90846	Hnr protein [impor
25	73.5	6.8	337	2 H85703	Hnr protein [impor
26	73.5	6.8	337	2 A36871	37K regulator resp
27	73.5	6.8	483	1 DESHGC	phosphogluconate d
28	73.5	6.8	727	2 AH0479	1,4-alpha-glucan b
29	72.5	6.7	191	2 S57642	interferon precurs

30	72.5	6.7	446	2 F71069	hypothetical prote
31	72.5	6.7	663	2 T03217	LIM domain protein
32	72	6.6	196	2 T01954	hypothetical prote
33	72	6.6	312	2 C84608	hypothetical prote
34	72	6.6	366	2 I53035	trithorax homolog
35	72	6.6	433	2 T43243	probable translati
36	72	6.6	792	2 T23742	hypothetical prote
37	72	6.6	811	2 E71281	methionine-tRNA li
38	72	6.6	3859	2 A48205	All-1 protein tGTE
39	72	6.6	3988	2 A44265	trithorax homolog
40	71.5	6.6	505	2 S13330	fixL protein - Bra
41	71	6.5	478	4 TVHUBD	transforming prote
42	71	6.5	567	2 S10138	mcf2 protein - hum
43	71	6.5	648	2 T23621	hypothetical prote
44	71	6.5	718	2 S06068	myeloperoxidase (E
45	71	6.5	925	1 TVHUBD	transforming prote

ALIGNMENTS

RESULT 1

G01233
phorbolins I - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 19-Apr-2002
C:Accession: G01233
R:Madson, P.P.
submitted to the EMBL Data Library, December 1993
A:Reference number: G06330
A:Accession: G01233
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-116 <MAD>
A:Cross-references: EMBL:003891; NID:g436940; PIDN:AAA03706.1; PID:g436941
C:Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1

Query Match 22.6%; Score 245; DB 2; Length 116;
Best Local Similarity 46.4%; Pred. No. 3.1e-17;
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

QY	76	YRVTFWTSWSPCYD--CARHVA	DLRGNP	LNLSRIETARLYFCEDRKAEP----	EGLRRLH 130
Db	7	YRVTFWTSWSPCYD--CARHVA	DLRGNP	LNLSRIETARLYFCEDRKAEP----	EGLRRLH 130
QY	131	RAGVQIAITMFKDYFCWNTFV	ENHERTFKAWGLHENSRLSRQLRRL	180	
Db	62	DAGAQVSIIMTYDEFYCWDTFV	RQGPCFQPDGLEDHQSALSGRLRAIL	111	

RESULT 2

A53853
apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 18-Aug-2000
C:Accession: A53853
R:Yamanaka, S.; Poksay, K.S.; Balestra, M.E.; Zeng, G.Q.; Innerarity, T.L.
J. Biol. Chem. 269, 21725-21734, 1994
A:Title: Cloning and mutagenesis of the rabbit Apob mRNA editing protein. A zinc mot
ributed.
A:Reference number: A53853; MUID:94342367; PMID:8063816
A:Accession: A53853
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-236 <YAM>
A:Cross-references: GB:U10695; NID:g506180; PIDN:AAA56718.1; PID:g506181
C:Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1
C:Keywords: hydrolase; zinc

Query Match 20.7%; Score 225; DB 2; Length 236;
Best Local Similarity 37.4%; Pred. No. 6.8e-15;
Matches 49; Conservative 26; Mismatches 48; Indels 8; Gaps 4;


```
QY 108 IFARLYFCEDRKAPEGLRLHRAGVOIALMTFKDYFYCWNTFVEN-----HERTKAW 163
Db 316 TLDGKAYIQD-----LSFKILDFEKTILENDTIYHNNPFIFYR 355
QY 164 GLHNSVRLSRLRRILLPIYEVDDL 190
Db 356 DLHE-----PLISIDDL 368

RESULT 10
hypoetical protein YGR198w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypoethical protein G7594
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
A:Accession: S53921; S53918; S64520; S64516
R:Nombela, C.
submitted to the EMBL Data Library, November 1994
A:Reference number: S53918
A:Accession: S53921
A:Molecule type: DNA
A:Residues: 1-817 <NOM>
A:Cross-references: EMBL:X82775; NID:g791126; PIDN:CAA58017.1; PID:g791130
R:Guereiro, P.; Silva, A.M.E.; Barreiros, T.; Arroyo, J.; Garcia-Gonzalez, M.; Garcia-S
Yeast 11, 1087-1091, 1995
A:Title: The complete sequence of a 9000 bp fragment of the right arm of Saccharomyces c
A:Reference number: S59180; MUID:96076633; PMID:7502584
A:Accession: S59183
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-817 <GUR>
A:Cross-references: EMBL:X82775; NID:g791126; PIDN:CAA58017.1; PID:g791130
R:Guereiro, P.; Barreiros, T.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64517
A:Accession: S64520
A:Molecule type: DNA
A:Residues: 1-817 <GUW>
A:Cross-references: EMBL:X82775; NID:g1323352; PIDN:CAA97225.1; PID:e243737; PID:g132335
R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64499
A:Accession: S64516
A:Molecule type: DNA
A:Residues: 1-240 <ARR>
A:Cross-references: EMBL:X82775; MIPS:YGR198w
C:Genetics:
A:Map position: 7R

Query Match 7.2%; Score 78.5; DB 2; Length 817;
Best Local Similarity 21.3%; Pred. No. 12;
Matches 42; Conservative 33; Mismatches 63; Indels 59; Gaps 9;
QY 41 SFSIDFGYLRNKNCHVELLFLRYISDWLDGRCYRVYTWFTSWSPCYDCARHVADEFLRG 100
Db 413 SESLD--WLENSTRC-----YYPGNISKVL-TNAWSTLYEIRKYQLDFLVS 455
QY 101 NPNLSLRIFTARYFCEDRK-----APEGLRL-----HRAGVOIAIMTF 141
Db 456 N-NLTSVLCNAMLSLTKERKNADVERQEEGEKALRELQFKYSYVTLAQORHIEITAKTL 514
QY 142 K-----DYFCWNTFV-----ENHERTEKAW-----EGLHNSVRLSRLRRI 179
Db 515 ESILSKNPNYKAWHLLALCRSVQEDKMSYKIVCSVLEAMNESLQNNTLLNDRWQFI 574
QY 180 LLPLYEVDLRLDAFRL 196
Db 575 HLKLTALALIEETFGTL 591

us-09-966-880a-8.rpr
QY 108 IFARLYFCEDRKAPEGLRLHRAGVOIALMTFKDYFYCWNTFVEN-----HERTKAW 163
Db 316 TLDGKAYIQD-----LSFKILDFEKTILENDTIYHNNPFIFYR 355
QY 164 GLHNSVRLSRLRRILLPIYEVDDL 190
Db 356 DLHE-----PLISIDDL 368

RESULT 11
hypoetical protein F31F4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
A:Accession: T32318
R:Blanchard, M.; Kramer, J.; Elliott, G.; Twyman, B.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid F31F4.
A:Reference number: Z21149
A:Accession: T32318
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-251 <BLA>
A:Cross-references: EMBL:AF024503; PIDN:AAB70382.1; GSPDB:GN00023; CBSP:F31F4.2
A:Experimental source: strain Bristol N2; clone F31F4
C:Genetics:
A:Gene: CBSP:F31F4.2
A:Map position: 5
A:Introns: 53/2
C:Superfamily: Caenorhabditis elegans hypothetical protein K0684.9

Query Match 7.1%; Score 77.5; DB 2; Length 251;
Best Local Similarity 26.8%; Pred. No. 3.9;
Matches 40; Conservative 13; Mismatches 51; Indels 45; Gaps 7;
QY 1 MDSLMLNRRKFLYQFKNVRWAKGRRETYLCYVVRDSDATSFSLDFG-----YLRN 51
Db 78 MLSLLMALNRRFAVF-----W-----YLRNRTFSNNTFMKFGVLFAVIYIYLD 126
QY 52 KNGCHVELLFLRY-----ISDWLDGRCYRVYTWFTSWSPCYDCARHVADEFLRGPNLSLR 107
Db 127 YLGC-----FPYEEKSYTFWLVDTPLCYSVSNY-----SDILFNN---SLV 165
QY 108 IFARLYFCEDRKAPEGLRLHRAGVOI 136
Db 166 VFTLAINLITVYKAGSNRNLQAAGOM 194

us-09-966-880a-8.rpr
QY 108 IFARLYFCEDRKAPEGLRLHRAGVOI 136
Db 166 VFTLAINLITVYKAGSNRNLQAAGOM 194

RESULT 12
hypoetical protein F56B6.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
A:Accession: T30999
R:Stellies, L.; Stellie, L.
submitted to the EMBL Data Library, September 1999
A:Description: The sequence of C. elegans cosmid F56B6.
A:Reference number: Z20957
A:Accession: T30999
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-328 <STE>
A:Cross-references: EMBL:U64599; PIDN:AAB04560.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: X
A:Introns: 10/2; 64/1; 93/2; 122/3; 155/2; 221/2; 285/2
A:Note: F56B6.5
C:Superfamily: vertebrate rhodopsin

Query Match 7.1%; Score 77.5; DB 2; Length 328;
Best Local Similarity 21.6%; Pred. No. 5.2;
Matches 41; Conservative 23; Mismatches 47; Indels 79; Gaps 11;
QY 49 LRNKGCHVELLFLRYISDWLDGRCYRVYTWFTSWSPC-----YDCARHVA 95
Db 127 MRPKNGTDSIENNNHLCIAHWSPSP-----HAQWI---SVCSVMIFILPGLVIFYCVYHV- 179
QY 96 DFLRGPNLSLRIFTARYFCEDRKAPEGLRLHR-----AGVOIAIMTFKD 143
```

Db 180 -----FCKLREA-AKGRRLHRNKRSSQRTSRVQVLF--- 216
QY 144 YFYCWNTFVFNHRTFKAWGLHNSVRLSKQLRRLPL- 216
Db 217 HLLCWSPF-----WLENLFAIFR--VRITQLMRIIIVHILPPYNCALNPVLYAYRA 269
QY 187 DLLRDAPRTL 196
Db 270 ENERTAFKSL 279

RESULT 13
VGHE50
glycoprotein D precursor - suid herpesvirus 1 (strain Rice)
C:Species: suid herpesvirus 1
A:Note: host Sus scrofa domestica (domestic pig)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
C:Accession: A27788
R:Petrovskis, E.A.; Timmins, J.G.; Armentrout, M.A.; Marchiolli, C.C.; Yancey Jr., R.J.;
J. Virol. 59, 216-223, 1986
A:Title: DNA sequence of the gene for pseudorabies virus gp50, a glycoprotein without N-
A:Reference number: A27788; MUID:86281819; PMID:3016293
A:Accession: A27788
A:Molecule type: DNA
A:Residues: 1-402 <PET>
A:Cross-references: GB:M14001; NID:g334051; PIDN:AAC35203.1; PID:g334052
C:Superfamily: herpesvirus glycoprotein D
C:Keywords: glycoprotein; transmembrane protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-402/Product: glycoprotein D #status predicted <GPD>
F:362-378/Domain: transmembrane #status predicted <TMN>

Query Match 7.1%; Score 77.5; DB 1; Length 402;
Best Local Similarity 24.9%; Pred. No. 6.6;
Matches 42; Conservative 17; Mismatches 73; Indels 37; Gaps 6;
QY 24 RRTYLCYVYVVRDSATSFSLDFGLYLNKNGCHVELLFLRYISDWLDP---GRCYRV 79
Db 88 RRTYRAHV-----AWYRIADGCAHLLYFIEVA---DCDPRQVFGRCRRRT 130
QY 80 WFTSWSPCYDCARHVADFLRGNNPLSLRIETARLYFCEDRKAEPEGLRLRHAGVQIA-- 137
Db 131 TPMWTPSADYMPTEDEL-----GLLWVAPGRFNEGQYRLVSDGVNLTLD 178
QY 138 -IMTFKDYFYCWNTFVFNHRTFKAWGLHNSVRLSKQLRRLPLLYE 185
Db 179 FMVALPEGQCPARVDQH-RTYKFGACWSDDSFKRGVDVWRFLTPFIQ 226

RESULT 14
G59432
MacGAP protein [imported] - huam
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 02-Aug-2002
C:Accession: G59432
R:Uchida, T.; Kuramasu, A.; Okumura, K.; Nakao, A.; Ogawa, H.; Ra, C.
submitted to GenBank, December 2001
A:Description: Molecular Cloning and Characterization of A novel GTPase Activating Prote
A:Reference number: G59432
A:Accession: G59432
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-618 <UCH>
A:Cross-references: GB:NP_277050; PID:g15723376; PIDN:NP_277050.1

Query Match 7.1%; Score 77.5; DB 2; Length 618;
Best Local Similarity 22.5%; Pred. No. 11;
Matches 42; Conservative 36; Mismatches 94; Indels 15; Gaps 6;
QY 16 KNRWAKGRRETYLCYVVKRRDSATSFSLDFGLYLNKNGCHVELLFLRYISD---WDLDP 72
Db 204 ETKQSKGDDATLPSFLPKDKTGTTRIGDLAPQDMKKVKCHLALIELTALYDVLGIELQ 263

QY 73 GRCYRVFTWFTSWSPCYDCARHVADFLRGNNPLSL-RIETARLYFCEDRKAEPEGLRLHR 131
Db 264 QKAVKIKTKDSGLFCVPLTALLEQDQKRVPGMRIPLIQKLLISRIEERGLETEGLLRIPG 323
QY 132 AGVQIAIMTFKDYFYCWNTFVFNHRTFKAWGL--HNSVRLSKQLRRLPLLYEVDL 189
Db 324 AAIRIKNL-----CQLEAKFYEGTFN-WESVQKHDAASLLKLFIRELPOLLSVEYL 375
QY 190 RDAFRTL 196
Db 376 K-AFAQV 381

RESULT 15
T14280
RWL protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14280
R:Wilkinson, R.; Fitter, S.; Tscharke, D.; Simmons, A.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z17958
A:Accession: T14280
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1829 <WIL>
A:Cross-references: EMBL:AF060565; NID:g3091277; PID:g3091278; PIDN:AAC15232.1
A:Experimental source: strain BALB/c; brain

Query Match 7.1%; Score 77; DB 2; Length 1829;
Best Local Similarity 27.2%; Pred. No. 40;
Matches 28; Conservative 15; Mismatches 26; Indels 34; Gaps 6;
QY 69 DLDPGRCYRVFT--WFTSWSPCYDCARHVADFLRGNNPLSLRIETARLYFCEDRKAEPEGL 126
Db 678 DLEPGKSKIANIYFDPGLQCGD-HRYI-----GLPFLLS-----KSEPK-- 715
QY 127 RRLHRAGVQIAIMTFKDYFYCWNTFVFNHRTFKAWGLHNS 169
Db 716 -----VQPGVAMQED---LWNADWDAAHQSLEFKAWMGIKENA 748

Search completed: June 14, 2003, 18:11:56
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 17:59:53 ; Search time 24 Seconds
(without alignments)
342.180 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLMMRRKFLQFKNVRW.....ILLPLYEVDDLRAFRITGL 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	364	33.5	382	1 PHB3_HUMAN	Q9uh17 homo sapien
2	363.5	33.5	199	1 PHB1_HUMAN	P31941 homo sapien
3	339	31.2	190	1 PHB2_HUMAN	Q9ue74 homo sapien
4	225	20.7	236	1 ABME_RABIT	P47855 oryctolaqus
5	216.5	19.9	236	1 ABME_HUMAN	P41238 homo sapien
6	211	19.4	229	1 ABME_MOUSE	P51908 mus musculus
7	198	18.2	229	1 ABME_RAT	P38483 rattus norv
8	97.5	9.0	369	1 CT19_YEAST	Q02732 saccharomyc
9	85	7.8	503	1 MATK_KUNBA	Q9tkc0 kunzea baxt
10	78.5	7.2	817	1 YG4B_YEAST	P46951 saccharomyc
11	77.5	7.1	402	1 VGLD_PVRRI	P07645 pseudorabie
12	77	7.1	1829	1 RW1_MOUSE	O70472 mus musculus
13	75.5	7.0	382	1 HUGA_APIME	Q08169 apis mellif
14	75	6.9	354	1 WN14_CHICK	O42280 gallus gall
15	73.5	6.8	337	1 HNR_ECOLI	P37055 escherichia
16	73.5	6.8	482	1 6PGD_SHEEP	P00349 ovis aries
17	73.5	6.8	738	1 TRFM_MOUSE	Q9r0r1 mus musculus
18	72.5	6.7	191	1 INF_ANAPL	P51526 anas platyr
19	72.5	6.7	641	1 PAPB_MOUSE	Q9wvp6 mus musculus
20	72	6.6	357	1 WN15_HUMAN	O14905 homo sapien
21	72	6.6	433	1 ERF1_SCHPO	P79063 schizosacch
22	72	6.6	792	1 GUAA_CAEEL	Q09580 caenorhabdi
23	72	6.6	811	1 SYM_TREPA	O83776 treponema p
24	72	6.6	1094	1 DRQL_GPCMV	Q69025 guinea pig
25	72	6.6	3866	1 HRX_MOUSE	P55200 mus musculus
26	72	6.6	3969	1 HRX_HUMAN	Q03164 homo sapien
27	71.5	6.6	505	1 PHTL_BRAJA	P23222 bradyrhizob
28	71	6.5	330	1 PXTD_COMTE	Q59727 comamonas t
29	71	6.5	718	1 PERM_MOUSE	P11247 mus musculus
30	71	6.5	925	1 DBL_HUMAN	P10911 homo sapien
31	71	6.5	1805	1 RW1_HUMAN	Q92545 homo sapien
32	70.5	6.5	630	1 F1MB_CHICK	P19179 gallus gall
33	70	6.4	626	1 HCY6_ANDAU	P80476 androctonus

RESULT 1

ID	PHB3_HUMAN	STANDARD;	PRT;	382 AA.
AC	Q9uh17: Q95618:			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Phorbolin 3 (APOBEC1-like).			
GN	APOBEC1L.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,			
RA	Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,			
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,			
RA	Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,			
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,			
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,			
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,			
RA	Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,			
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,			
RA	Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,			
RA	Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,			
RA	Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,			
RA	Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,			
RA	Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,			
RA	McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,			
RA	Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,			
RA	Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,			
RA	Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,			
RA	Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,			
RA	Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,			
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,			
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,			
RA	Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,			
RA	Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuayama S.,			
RA	Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,			
RA	Dorman A., Fang F., Fu Y., Hua A., Kenton S., Lai H., Lao H.I.,			
RA	Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,			
RA	Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,			
RA	Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,			
RA	Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,			
RA	Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,			
RA	Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,			
RA	Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,			
RA	Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,			
RA	Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,			
RA	Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,			
RA	Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,			
RA	Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,			
RA	Seroussi E., Franconi I., Tapia I., Bruder C.E., O'Brien K.P.,			
RA	Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,			

ALIGNMENTS

P10844 clostridium
P75919 escherichia
Q9nrj5 homo sapien
Q9z9h0 pseudomonas
P57328 buchnera ap
P29416 mus musculu
Q91637 xenopus lae
P51565 escherichia
P13561 red clover
P29837 l genome po
P11935 c cephalosp
P55615 rhizobium s

34 70 6.4 1290 1 BXB_CLOBO
35 69.5 6.4 493 1 YMDC_ECOLI
36 69.5 6.4 636 1 PAPB_HUMAN
37 69 6.4 900 1 GLND_PSEAE
38 69 6.4 1070 1 EX5C_BUCAI
39 68.5 6.3 528 1 HEXA_MOUSE
40 68.5 6.3 682 1 KIF2_XENLA
41 68.5 6.3 988 1 TNP9_ECOLI
42 68.5 6.3 996 1 VGNM_CMV
43 68.5 6.3 3414 1 POLG_LANVT
44 68 6.3 332 1 EXPA_CEPAC
45 68 6.3 387 1 Y4PF_RHISN

```

RA Ttlahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
[2]
RP SEQUENCE OF 148-382 FROM N.A.
RC TISSUE-Keratinocytes;
RA Madsen P.;
RT "Molecular cloning of phorbol 1.3.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-148 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL022318; CAB45270.1; -.
DR EMBL; U61084; AAD00090.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 2.
KW Hydrolase.
SQ SEQUENCE 382 AA; 45924 MW; DA6EDD23E8856240 CRC64;

Query Match 33.5%; Score 364; DB 1; Length 382;
Best Local Similarity 43.5%; Pred. No. 2.5e-29;
Matches 83; Conservative 26; Mismatches 62; Indels 20; Gaps 6;

QY 5 LMNRKFLYQFKNVWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRN--KN-----G 54
Db 192 LMDPDTFTFNNDPLVLRRTYLCYEVEERLDNGTWLMDQHWGFLCNKLLCGFYG 251
QY 55 CHVELLFLRYISDWLDPGRCYRVTFWTSWSPCYD--CARHVADFLRGNPNLSLRIFTAR 112
Db 252 RHAEFLDLVPSQLDPAQIYRVTFWISWPCFSWGCAGEVRAFLQENTHVRLEIFAAR 311
QY 113 LYCEDRKAEP---EGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGLEHNS 169
Db 312 IY-----DYDPLYKEALQMLRDAGAQSIMTYDEFCWNTFVYRGCPFPQWDGLEHS 366
QY 170 VRLSRLRL 180
Db 367 QALSGRLRAIL 377

RESULT 2
PHBL_HUMAN
ID PHBL_HUMAN STANDARD; PRT; 199 AA.
AC P31941; Q12807;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phorbol 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Epidermis;
RX MEDLINE=99399284; PubMed=10469298;
RA Madsen P.P., Anant S., Rasmussen H.H., Gromov P., Vorum H.,
RA Dumanski J.P., Tommerup N., Collins J.E., Wright C.L., Dunham I.,
RA Macginnitie A.J., Davidson N.O., Celis J.E.;
RT "psoriasis up-regulated phorbol-1 shares structural but not
RT functional similarity to the mRNA-editing protein apobec-1.";
RL J. Invest. Dermatol. 113:162-169(1999).
[2]

```

```

RP SEQUENCE OF 53-60; 112-121 AND 129-137.
RC TISSUE-Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03891; AAA03706.2; -.
DR Aarhus/Ghent-2DPAGE; 2116; IEF.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.
KW Hydrolase.
SQ SEQUENCE 199 AA; 23012 MW; 42E99E0D7DF7AA14 CRC64;

Query Match 33.5%; Score 363.5; DB 1; Length 199;
Best Local Similarity 43.5%; Pred. No. 1.3e-29;
Matches 83; Conservative 28; Mismatches 57; Indels 23; Gaps 7;

QY 5 LMNRKFLYQFKNVWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRN--KN-----G 54
Db 12 LMDPHFTSFNFN---GIGRHKTYLCYEVEERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG 68
QY 55 CHVELLFLRYISDWLDPGRCYRVTFWTSWSPCYD--CARHVADFLRGNPNLSLRIFTAR 112
Db 69 RHAEFLDLVPSQLDPAQIYRVTFWISWPCFSWGCAGEVRAFLQENTHVRLEIFAAR 128
QY 113 LYCEDRKAEP---EGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGLEHNS 169
Db 129 IY-----DYDPLYKEALQMLRDAGAQSIMTYDEFCWNTFVYRGCPFPQWDGLEHS 183
QY 170 VRLSRLRL 180
Db 184 QALSGRLRAIL 194

RESULT 3
PHB2_HUMAN
ID PHB2_HUMAN STANDARD; PRT; 190 AA.
AC Q9UE74;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phorbol 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Keratinocytes;
RA Madsen P.;
RT "Molecular cloning of phorbol 2.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; U61083; AAD00089.1; --
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; FALSE_NEG.
KW Hydrolase.
SQ SEQUENCE 190 AA; 22453 MW; A54DCBC100FC26F3 CRC64;

Query Match 31.28; Score 339; DB 1; Length 190;
Best Local Similarity 41.18; Pred. No. 3.7e-27;
Matches 78; Conservative 27; Mismatches 65; Indels 20; Gaps 6;

QY 6 MNRKFLYQFNWAKGRRTYLYCYVVKRDSATSFSLD--FGYLRN--KN-----GC 55
DB 1 MDPPTFFNFNDPLVLRRTGYLCYEYERLDNGWVLLDQHMGLCNKLLCGFYGR 60
QY 56 HVELLFYRISDMDLDFRCYRVTVFTSWSPCYD--CARHVADEFLRGNPNLSLRIFATRL 113
DB 61 HADNREFDLVPSLQDPAQIVRVTVFTSWSPCFESWGCAGEVRAFLQENTHVRPLPIFAARI 120
QY 114 YFCEDRAEP---EGLRLRHAGVQIAIMTFKDYFCWNTFVEVHERTFKAWEGHNSV 170
DB 121 Y-----DYDPLYKEALQMLRDAGAQSIMTYDEFCYWDTEVYRQCGFPQPDWGLEHSQ 175
QY 171 RLSRLRLRL 180
DB 176 ALSRLRLAIL 185

RESULT 4
ABME_RABBIT
ID ABME_RABBIT STANDARD; PRT; 236 AA.
AC P47855;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Apolipoprotein B mRNA editing protein (APOBEC-1).
GN APOBEC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC STRAIN-New Zealand white; TISSUE=Small intestine;
RX MEDLINE=94342367; PubMed=8063816;
RA Yamanaka S., Poksay K.S., Balestra M.E., Zeng G.-Q., Innerarity T.L.;
RT "Cloning and mutagenesis of the rabbit Apob mRNA editing protein. A
RT zinc motif is essential for catalytic activity, and noncatalytic
RT auxiliary factor(s) of the editing complex are widely distributed";
RL J. Biol. Chem. 269:21725-21734(1994).
CC -!- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC -!- COFACTOR: ZINC.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; U10695; AAA56718.1; --
DR InterPro; IPR002125; dCMP/cyt_deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.
KW mRNA processing; Hydrolase; Zinc.

FT METAL 61 ZINC (BY SIMILARITY).
FT METAL 93 ZINC (BY SIMILARITY).
FT METAL 96 ZINC (BY SIMILARITY).
FT DOMAIN 180 LEU-RICH MOTIF.
FT MUTAGEN 61 H->A: NONE OR LITTLE EDITING ACTIVITY.
FT MUTAGEN 61 H->C: RETAINS MOST EDITING ACTIVITY.
FT MUTAGEN 63 E->A: NONE OR LITTLE EDITING ACTIVITY.
FT MUTAGEN 92 P->A: RETAINS MOST EDITING ACTIVITY.
FT MUTAGEN 93 C->A: NONE OR LITTLE EDITING ACTIVITY.
FT MUTAGEN 96 C->A: NONE OR LITTLE EDITING ACTIVITY.
SQ SEQUENCE 236 AA; 27719 MW; AB3041CA5102F1F3 CRC64;

Query Match 20.7%; Score 225; DB 1; Length 236;
Best Local Similarity 37.4%; Pred. No. 1.5e-15;
Matches 49; Conservative 26; Mismatches 48; Indels 8; Gaps 4;

QY 24 RRETYLCYVVKRRDSATSFSLDFGYLRNKGCG-HVELLF-LRYSMDLDLPGRCYRVTVWF 81
DB 33 RKEACLLYEIKWGASSKTRSS-----GKNTTNHVEVNFLEKLTSEGLGPGSTCCSITWF 87
QY 82 TSWSPCYDCARHVADEFLRGNPNLSLRIFATRYFCEDRAEPGLRLRHAGVQIAIMTF 141
DB 88 LSWSPCFESWMAIREFLSQHPGVTLIFVARLFQHMRRR-NRQGLKDLVTSGTVKVMVS 146
QY 142 KDIFYCWNTFV 152
DB 147 SEYCYCWENFV 157

RESULT 5
ABME_HUMAN
ID ABME_HUMAN STANDARD; PRT; 236 AA.
AC P41238;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Apolipoprotein B mRNA editing protein (APOBEC-1) (HEPR).
GN APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94268910; PubMed=8208612;
RA Hadjiagapiou C., Giannoni F., Funahashi T., Skarosi S.F.,
RT Davidson N.O.;
RT "Molecular cloning of a human small intestinal apolipoprotein B mRNA
RT editing protein";
RL Nucleic Acids Res. 22:1874-1879(1994).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Small intestine;
RX MEDLINE=94359963; PubMed=8078915;
RA Lau P., Zhu H.-J., Baldini A., Charnsangavej C., Chan L.;
RT "Dimeric structure of a human apolipoprotein B mRNA editing protein
RT and cloning and chromosomal localization of its gene";
RL Proc. Natl. Acad. Sci. U.S.A. 91:8522-8526(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood leukocytes;
RX MEDLINE=98140126; PubMed=9479499;
RA Fujino T., Navaratnam N., Scott J.;
RT "Human apolipoprotein B RNA editing deaminase gene (APOBEC1).";
RL Genomics 47:266-275(1998).
CC -!- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF184722; AAF05929.1; .
DR InterPro; IPR000442; Intron_maturase2.
DR Pfam; PF01348; Intron_maturas2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 503 AA; 60203 MW; B114A4704FCEA059 CRC64;

Query Match 7.8%; Score 85; DB 1; Length 503;
Best Local Similarity 25.0%; Pred. No. 0.47; Mismatches 52; Indels 30; Gaps 8;
Matches 35; Conservative 23;
QY 31 YVVKRRDSATSFSLD-----FGYLRNK-----NGCHVELLF--LRVSDWDL 70
DB 112 FSLERKEIAKSYNLSRHSIFSLEDFKFTHLDYVSDVLPYHILHLEILXOTLRY---WVK 168
QY 71 DPGRCYRVWTFTS--WSPCYDCARHVAFLRGPNLSLRIPTARLYFCEDRKAEPGLRR 128
DB 169 DASLHLRLFLHDYNSFTTPKKHITFFLKGNPRLFLYNS--HICEYEYIFP-FLRN 225
QY 129 --LHRAGVQIAIMTFKDYFY 146
DB 226 QSSHRLSTSSGIFXERYFY 245

RESULT 10
YQ4B_YEAST STANDARD; PRT; 817 AA.
AC P46951;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 95.4 kDa protein in SNGI-PMT6 intergenic region.
GN YGR198W OR G7594.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=96076633; PubMed=7502584;
RA Guerreiro P., Maia e Silva A., Barreiros T., Arroyo J.,
RA Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C.,
RA Nombela C.;
RT "The complete sequence of a 9000 bp fragment of the right arm of
RT Saccharomyces cerevisiae chromosome VII contains four previously
RT unknown open reading frames.";
RL Yeast 11:1087-1091(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X82775; CAA58017.1; .
DR EMBL; 272983; CAA97225.1; .
DR SGD; S0003430; YGR198W.
KW Hypothetical protein.
SQ SEQUENCE 817 AA; 95365 MW; AC42730C8B9C3E4B CRC64;

Query Match 7.2%; Score 78.5; DB 1; Length 817;
Best Local Similarity 25.0%; Pred. No. 0.47; Mismatches 52; Indels 30; Gaps 8;
Matches 35; Conservative 23;
QY 31 YVVKRRDSATSFSLD-----FGYLRNK-----NGCHVELLF--LRVSDWDL 70
DB 112 FSLERKEIAKSYNLSRHSIFSLEDFKFTHLDYVSDVLPYHILHLEILXOTLRY---WVK 168
QY 71 DPGRCYRVWTFTS--WSPCYDCARHVAFLRGPNLSLRIPTARLYFCEDRKAEPGLRR 128
DB 169 DASLHLRLFLHDYNSFTTPKKHITFFLKGNPRLFLYNS--HICEYEYIFP-FLRN 225
QY 129 --LHRAGVQIAIMTFKDYFY 146
DB 226 QSSHRLSTSSGIFXERYFY 245

RESULT 10
YQ4B_YEAST STANDARD; PRT; 817 AA.
AC P46951;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 95.4 kDa protein in SNGI-PMT6 intergenic region.
GN YGR198W OR G7594.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=96076633; PubMed=7502584;
RA Guerreiro P., Maia e Silva A., Barreiros T., Arroyo J.,
RA Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C.,
RA Nombela C.;
RT "The complete sequence of a 9000 bp fragment of the right arm of
RT Saccharomyces cerevisiae chromosome VII contains four previously
RT unknown open reading frames.";
RL Yeast 11:1087-1091(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X82775; CAA58017.1; .
DR EMBL; 272983; CAA97225.1; .
DR SGD; S0003430; YGR198W.
KW Hypothetical protein.
SQ SEQUENCE 817 AA; 95365 MW; AC42730C8B9C3E4B CRC64;

Query Match 7.2%; Score 78.5; DB 1; Length 817;
Best Local Similarity 25.0%; Pred. No. 0.47; Mismatches 52; Indels 30; Gaps 8;
Matches 35; Conservative 23;
QY 31 YVVKRRDSATSFSLD-----FGYLRNK-----NGCHVELLF--LRVSDWDL 70
DB 112 FSLERKEIAKSYNLSRHSIFSLEDFKFTHLDYVSDVLPYHILHLEILXOTLRY---WVK 168
QY 71 DPGRCYRVWTFTS--WSPCYDCARHVAFLRGPNLSLRIPTARLYFCEDRKAEPGLRR 128
DB 169 DASLHLRLFLHDYNSFTTPKKHITFFLKGNPRLFLYNS--HICEYEYIFP-FLRN 225
QY 129 --LHRAGVQIAIMTFKDYFY 146
DB 226 QSSHRLSTSSGIFXERYFY 245

RESULT 10
YQ4B_YEAST STANDARD; PRT; 817 AA.
AC P46951;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 95.4 kDa protein in SNGI-PMT6 intergenic region.
GN YGR198W OR G7594.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=96076633; PubMed=7502584;
RA Guerreiro P., Maia e Silva A., Barreiros T., Arroyo J.,
RA Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C.,
RA Nombela C.;
RT "The complete sequence of a 9000 bp fragment of the right arm of
RT Saccharomyces cerevisiae chromosome VII contains four previously
RT unknown open reading frames.";
RL Yeast 11:1087-1091(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X82775; CAA58017.1; .
DR EMBL; 272983; CAA97225.1; .
DR SGD; S0003430; YGR198W.
KW Hypothetical protein.
SQ SEQUENCE 817 AA; 95365 MW; AC42730C8B9C3E4B CRC64;

Query Match 7.2%; Score 78.5; DB 1; Length 817;
Best Local Similarity 25.0%; Pred. No. 0.47; Mismatches 52; Indels 30; Gaps 8;
Matches 35; Conservative 23;
QY 31 YVVKRRDSATSFSLD-----FGYLRNK-----NGCHVELLF--LRVSDWDL 70
DB 112 FSLERKEIAKSYNLSRHSIFSLEDFKFTHLDYVSDVLPYHILHLEILXOTLRY---WVK 168
QY 71 DPGRCYRVWTFTS--WSPCYDCARHVAFLRGPNLSLRIPTARLYFCEDRKAEPGLRR 128
DB 169 DASLHLRLFLHDYNSFTTPKKHITFFLKGNPRLFLYNS--HICEYEYIFP-FLRN 225
QY 129 --LHRAGVQIAIMTFKDYFY 146
DB 226 QSSHRLSTSSGIFXERYFY 245

Best Local Similarity 21.3%; Pred. No. 3.7;
Matches 42; Conservative 33; Mismatches 63; Indels 59; Gaps 9;
QY 41 SFSLDGFLRNKNGCHVELLFRLYISDWDLDPGRCYRVWTFTSWSPCYDCARHVAFLRG 100
DB 413 SESLD--WLENSTRC-----VYPNISKVL-TNAWSTLYEIRKYOLDPLVS 455
QY 101 NPNSLRIPTARLYFCEDRK-----AEPEGLRR-----HRAGVQIAIMTF 141
DB 456 N-NLTSYLCNAMMLSTKEKNADVDEGECEKEKALRELQFKYSYTLAQORHIEIAKTL 514
QY 142 K-----DYFYCWNTFV-----ENHRTFAW-----EGLHNSVRLSRRLR 179
DB 515 ESILSKNPNTYKAWHLLALCRVSQEDKEMSYKIVCSVLEAMNESLQNTLLNDRWQFI 574
QY 180 LLPLYEVDLRLDAFRTL 196
DB 575 HLKLTQLALIEEIFGTL 591

RESULT 11
VGLD_PVRI STANDARD; PRT; 402 AA.
AC P07645;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein gp50.
OS Pseudorabies virus (strain Rice) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86281819; PubMed=3016293;
RA Petrovskis E.A., Timmins J.G., Armentrout M.A., Marchiolli C.C.,
RA Yancey R.J. Jr., Post L.E.;
RT "DNA sequence of the gene for pseudorabies virus gp50, a glycoprotein
RT without N-linked glycosylation.";
RL J. Virol. 59:216-223(1986)
CC - MISCELLANEOUS: GLYCOPROTEIN GP50 DOES NOT CONTAIN N-LINKED
CC CARBOHYDRATE, AS PREDICTED FROM ITS SEQUENCE.
CC - SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M14001; AAC35203.1; .
DR PIR; A27788; VGBE50.
DR InterPro; IPR002896; Herpes_glycop_D.
DR Pfam; PF01537; Herpes_glycop_D; 1.
KW Glycoprotein.
SQ SEQUENCE 402 AA; 44501 MW; B8763305995871E8 CRC64;

Query Match 7.1%; Score 77.5; DB 1; Length 402;
Best Local Similarity 24.9%; Pred. No. 2;
Matches 42; Conservative 17; Mismatches 73; Indels 37; Gaps 6;
QY 24 RRETYLYCVVKRRDSATSFSLDFGLRNKNGCHVELLFRLYISDWDLDP-----GRCYRV 79
DB 88 RRETYRAHV-----AWYRIADGCAHLLYFIEVA---DCDPQVFGRCRRRT 130
QY 80 WFTSWSPCYDCARHVAFLRGPNLSLRIPTARLYFCEDRKAEPGLRLRLHRAVQIA-- 137
DB 131 TPMMWTSPADYMEPTDEL-----GULMVAPGRFNEGQVRLVSDGVNLT 178
QY 138 -IMTFKDYFYCWNTFVENHERTEKAWEGHNSVRLSRQLRLLPLYE 185

Query Match 7.1%; Score 77.5; DB 1; Length 402;
Best Local Similarity 24.9%; Pred. No. 2;
Matches 42; Conservative 17; Mismatches 73; Indels 37; Gaps 6;
QY 24 RRETYLYCVVKRRDSATSFSLDFGLRNKNGCHVELLFRLYISDWDLDP-----GRCYRV 79
DB 88 RRETYRAHV-----AWYRIADGCAHLLYFIEVA---DCDPQVFGRCRRRT 130
QY 80 WFTSWSPCYDCARHVAFLRGPNLSLRIPTARLYFCEDRKAEPGLRLRLHRAVQIA-- 137
DB 131 TPMMWTSPADYMEPTDEL-----GULMVAPGRFNEGQVRLVSDGVNLT 178
QY 138 -IMTFKDYFYCWNTFVENHERTEKAWEGHNSVRLSRQLRLLPLYE 185

[illegible]


```
-----NOT ANNOTATED_CDS.
CC  EMBL; M64675; -; NOT ANNOTATED_CDS.
DR  EMBL; X66003; CAA46802.1; -.
DR  EMBL; AE000222; AAC74317.1; -.
DR  EMBL; D90758; BAA36103.1; -.
DR  EMBL; D90759; BAA36115.1; -.
DR  EMBL; D90852; BAA16029.1; -.
DR  EMBL; AE005341; AAG56092.1; -.
DR  EMBL; AP002556; BAB35160.1; -.
DR  PIR; S28505; S28505
DR  PIR; A36871; A36871.
DR  HSP; P52934; IQMP.
DR  EcoGene; EG12121; hnr.
DR  InterPro; IPR001789; Response_reg.
DR  Pfam; PF00072; response_reg.1.
DR  ProDom; PD000039; Response_reg; 1.
DR  SMART; SM00448; REC.1.
DR  PROSITE; PS0110; RESPONSE_REGULATORY; 1.
KW  Phosphorylation; Transcription regulation; DNA-binding;
KW  Sensory transduction; Complete proteome.
FT  DOMAIN 9 123 RESPONSE_REGULATORY.
FT  MOD_RES 58 58 PHOSPHORYLATION (BY SIMILARITY).
SQ  SEQUENCE 337 AA; 37302 MW; AB962EF94BC7B470 CRC64;

Query Match      6.8%; Score 73.5; DB 1; Length 337;
Best Local Similarity 21.6%; Pred. No. 4.2;
Matches 32; Conservative 19; Mismatches 52; Indels 45; Gaps 5;

Qy  79 TWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPEGRLRLH----- 130
Db  :||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
25 SWFSSLGATTVLAADGVDALE-----LLGGFTPDLMICDIAMPRMNGKLLLEHNRGDQ 79
Qy  131 -----RAGVO-IAIMTFKD-----YFYC-----WNTFVENHRT 158
Db  | | | : : : | | | | | | | | | | | | | | | | | | | | | |
80 TPVLVISITENMADIAKALRGVEDVLKPKVDLNRLREMFVACLYPSMFNSRVEEERL 139

Qy  159 FKAWEGHENSVRLSRQLRILLPLYEV 186
Db  | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
140 FRDWDAMVDNPAAAAAKLLQELQPPVQV 167
```

Search completed: June 14, 2003, 18:09:42
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:05:13 ; Search time 78 Seconds
(without alignments)
523.043 Million cell updates/sec

Title: US-09-966-880A-8
Perfect score: 1086
Sequence: 1 MDSLIMNRKFLYQKQNRW.....ILLPLYVDLDRFAFTLGL 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phase:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_virus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	198	4 Q9GZX7	Q9gzx7 homo sapien
2	1008	92.8	198	11 Q9WVE0	Q9wve0 mus musculu
3	390	35.9	384	4 Q9HCL6	Q9hcl6 homo sapien
4	369.5	34.0	204	4 Q9Y555	Q9y555 homo sapien
5	367.5	33.8	190	4 Q96F12	Q96f12 homo sapien
6	354	32.6	294	4 Q9NRW3	Q9nrw3 homo sapien
7	349	32.1	429	11 Q99J72	Q99j72 mus musculu
8	336.5	31.0	386	4 Q96AK3	Q96ak3 homo sapien
9	315.5	29.1	139	4 Q9Y4V1	Q9y4v1 homo sapien
10	278	25.6	224	11 Q9WV35	Q9wv35 mus musculu
11	274	25.2	224	4 Q9Y235	Q9y235 homo sapien
12	270.5	24.9	121	4 Q9Y553	Q9y553 homo sapien
13	257.5	23.7	103	4 Q9Y550	Q9y550 homo sapien
14	251	23.1	111	4 Q9UH18	Q9uh18 homo sapien
15	240.5	22.1	121	4 Q9Y554	Q9y554 homo sapien
16	238.5	22.0	131	4 Q9Y930	Q9y930 homo sapien

17	238.5	22.0	245	4 Q9Y549	Q9y549 homo sapien
18	219	20.2	229	11 Q9EQP0	Q9eqp0 mesocricetu
19	218.5	20.1	236	4 Q9UM71	Q9um71 homo sapien
20	213.5	19.7	235	6 Q9TUI7	Q9tui7 monodelphis
21	211	19.4	229	11 Q99L67	Q99l67 mus musculu
22	198	18.2	127	4 Q9H9H8	Q9h9h8 homo sapien
23	182.5	16.8	77	4 Q9Y551	Q9y551 homo sapien
24	179	16.5	132	4 Q9UE64	Q9ue64 homo sapien
25	102	9.4	720	5 Q9W293	Q9w293 drosophila
26	85	7.8	414	17 Q97L68	Q97l68 sulfolobus
27	84.5	7.8	504	8 Q9TKC5	Q9tkc5 asteromyrtu
28	83.5	7.7	504	8 Q9TKC4	Q9tkc4 asteromyrtu
29	83.5	7.7	533	11 Q70276	Q70276 rattus norv
30	82.5	7.6	483	11 Q8VHP2	Q8vhp2 mus musculu
31	82.5	7.6	533	11 Q91Z05	Q91zg5 mus musculu
32	81.5	7.5	400	12 Q9IW02	Q9iw02 pseudorabie
33	81.5	7.5	516	12 Q9IW00	Q9iw00 pseudorabie
34	81.5	7.5	2342	5 Q46048	Q46048 drosophila
35	80.5	7.4	505	8 Q9TKB9	Q9tkb9 kunzea capi
36	80	7.4	476	16 Q25366	Q25366 helicobacte
37	79.5	7.3	503	8 Q9TKC3	Q9tkc3 asteromyrtu
38	79.5	7.3	504	8 Q9MSA5	Q9msa5 podocarpus
39	78.5	7.2	503	8 Q9TKB1	Q9tkb1 leptospermu
40	78	7.2	504	8 Q95GT9	Q95gt9 nepenthes m
41	77.5	7.1	251	5 Q17138	Q17138 caenorhabdi
42	77.5	7.1	279	12 Q91F41	Q91f41 cydia pomon
43	77.5	7.1	369	5 Q22995	Q22995 caenorhabdi
44	77.5	7.1	404	12 Q92290	Q92290 pseudorabie
45	77.5	7.1	618	4 Q96S64	Q96s64 homo sapien

ALIGNMENTS

RESULT 1

Q9GZX7	PRELIMINARY;	PRT;	198 AA.
AC Q9GZX7;			
DT 01-MAR-2001 (TREMBlrel. 16, Created)			
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)			
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)			
DE Activation-induced cytidine deaminase.			
GN AID.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=20408890; PubMed=10950930;			
RA Muto T., Muranatsu M., Taniwaki M., Kinoshita K., Honjo T.;			
RT "Isolation, tissue distribution and chromosomal localization of the			
RL human activation-induced cytidine deaminase (hAID) gene."			
RL Genomics 68:85-88(2000).			
RN [2]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=20460541; PubMed=11007475;			
RA Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,			
RA Catalan N., Forveille M., Dufourcq-Lagelouse R., Gennery A.,			
RA Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N.,			
RA Muramatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A.,			
RA Durandy A.;			
RT "Activation-induced cytidine deaminase (AID) deficiency causes the			
RT autosomal recessive form of the Hyper-IgM syndrome (HIGM2).";			
RL Cell 102:565-575(2000).			
RN [3]			
RP SEQUENCE FROM N.A.			
RC TISSUE-B-CELL;			
RA Strausberg R.;			
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AB040431; BAB12721.1; -			
DR EMBL; AB040430; BAB12720.1; -			
DR EMBL; BC006296; AAH06296.1; -			

```
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 198 AA; 23953 MW; 3C27BB143DB184A9 CRC64;

Query Match 100.0%; Score 1086; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 7.4e-102; Indels 0; Gaps 0;
Matches 198; Conservative 0; Mismatches 0;

QY 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
DB 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60

QY 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNSLRIFTARLYFCEDRK 120
DB 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNSLRIFTARLYFCEDRK 120

QY 121 ASPEGLRLRHAGVOIAIMTKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSRLRRL 180
DB 121 ASPEGLRLRHAGVOIAIMTKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSRLRRL 180

QY 181 LPLYEVDDLRLDAFRITGL 198
DB 181 LPLYEVDDLRLDAFRITGL 198

RESULT 2
Q9WVE0 PRELIMINARY; PRT; 198 AA.
AC Q9WVE0;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Activation-induced cytidine deaminase.
GN AICDA OR AID.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303612; PubMed=10373455;
RA Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,
RA Davidson N.O., Honjo T.;
RT "Specific expression of activation-induced cytidine deaminase (AID), a
RT novel member of the RNA-editing deaminase family in germinal center B
RT cells.";
RL J. Biol. Chem. 274:18470-18476(1999).
DR EMBL; AF132979; AAD41793.1; -.
DR MGD; MGI:1342279; Aicda.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BEB2 CRC64;

Query Match 92.8%; Score 1008; DB 11; Length 198;
Best Local Similarity 92.9%; Pred. No. 5.5e-94;
Matches 183; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
DB 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60

QY 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNSLRIFTARLYFCEDRK 120
DB 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNSLRIFTARLYFCEDRK 120

QY 121 ASPEGLRLRHAGVOIAIMTKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSRLRRL 180
DB 121 ASPEGLRLRHAGVOIAIMTKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSRLRRL 180

QY 181 LPLYEVDDLRLDAFRITGL 197
DB 181 LPLYEVDDLRLDAFRITGL 197

RESULT 3
Q9HC16 PRELIMINARY; PRT; 384 AA.
AC Q9HC16;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE MDS019 (Phorbol-like protein MDS019).
GN MDS019.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.;
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT myelodysplastic Syndromes patient.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF182420; AAG14956.1; -.
DR EMBL; BC024268; AAH24268.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 384 AA; 46408 MW; 60525DC3B7D903D6 CRC64;

Query Match 35.9%; Score 390; DB 4; Length 384;
Best Local Similarity 44.9%; Pred. No. 3.1e-31;
Matches 83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;

QY 6 MNRKRFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRNK-----NGC 55
DB 197 MDPPTFTFNFPNPPWVRGRHETLYCYEYERHNDTWLLNQRRGFLCNQAPKHGFLRGR 256

QY 56 HVELLFLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNSLRIFTARLYF 115
DB 257 HAEICFLDVIPIFWKLLDQDQIRVTCFTSWSPCFCAQEMAKFIKNNKIVSLCIFTARIY- 315

QY 116 CEDRAKEPEGLRLRHAGVOIAIMTKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSRQ 175
DB 316 --DQGRCEGLRTLAEGAKISIMTYSEFKHCWDTFVDHGCPCFPQDPDGLDEHSQDLGR 374

QY 176 LRRIL 180
DB 375 LRAIL 379

RESULT 4
Q9Y555 PRELIMINARY; PRT; 204 AA.
AC Q9Y555;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE BK150C2.3 (Putative novel protein similar to APOBEC1 (Apolipoprotein B
DE mRNA editing protein) and Phorbol) (Fragment).
GN BK150C2.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45271.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
```

```
KW Lipoprotein.
FT NON_TER
SQ SEQUENCE 204 AA; 24285 MW; 79C656F580A40554 CRC64;

Query Match 34.0%; Score 369.5; DB 4; Length 204;
Best Local Similarity 44.9%; Pred. No. 1.8e-29;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
DB 31 FYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCHAEKRCFLSW 88

QY 65 ISDWDLPGRCYRVWTFTSWSPCYDCARHVADEFGRGNPNLSLRIFTARLYFCEDRKAEP 124
DB 89 FCDDILSPNTKYQVTWYTSWSPCPDCEGAEVAFELARHSNVNLTIFTARLYYFQ-YPCYQE 147

QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQLRILL 180
DB 148 GLRSLSQEGVAEIMDYEDFKYCWENFYVNDNEPFPKPKGLKTNFRLKRLRRESL 203

RESULT 5
Q96F12 PRELIMINARY; PRT; 190 AA.
AC Q96F12;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011739; AAH11739.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 190 AA; 22827 MW; DA0584EF75C91CF0 CRC64;

Query Match 33.8%; Score 367.5; DB 4; Length 190;
Best Local Similarity 44.9%; Pred. No. 2.6e-29;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
DB 17 FYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCHAEKRCFLSW 74

QY 65 ISDWDLPGRCYRVWTFTSWSPCYDCARHVADEFGRGNPNLSLRIFTARLYFCEDRKAEP 124
DB 75 FCDDILSPNTKYQVTWYTSWSPCPDCEGAEVAFELARHSNVNLTIFTARLYYFQ-YPCYQE 133

QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQLRILL 180
DB 134 GLRSLSQEGVAEIMDYEDFKYCWENFYVNDNEPFPKPKGLKTNFRLKRLRRESL 189

RESULT 6
Q9NRW3 PRELIMINARY; PRT; 294 AA.
AC Q9NRW3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Phorbolin I protein.
GN PBI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J., Huang Q., Yu Y., Xu S., Wang Y., Han Z., Chen Z., Zhou J.,
RA Tu Y., Gu W., Fu G., Huang C.;
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT Myelodysplastic Syndromes patient.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF165520; AAF86650.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 294 AA; 33363 MW; IB39C7A13D690901 CRC64;

Query Match 32.6%; Score 354; DB 4; Length 294;
Best Local Similarity 43.9%; Pred. No. 9.8e-28;
Matches 75; Conservative 26; Mismatches 60; Indels 10; Gaps 5;

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
DB 17 FYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCHAEKRCFLSW 74

QY 65 ISDWDLPGRCYRVWTFTSWSPCYDCARHVADEFGRGNPNLSLRIFTARLYFCEDRKAEP 124
DB 75 FCDDILSPNTKYQVTWYTSWSPCPDCEGAEVAFELARHSNVNLTIFTARLYYFQ-YPCYQE 133

QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQL 175
DB 134 GLRSLSQEGVAEIMDYEDFKYCWENFYVNDNEPFPKPKWGI-KNOLSTSEK 183

RESULT 7
Q99J72 PRELIMINARY; PRT; 429 AA.
AC Q99J72;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 51.0 kDa protein.
GN BC003314.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003314; AAH03314.1; -.
DR MGD; MGI:1933111; BC003314.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 51017 MW; BE44D01380AD7F6E CRC64;

Query Match 32.1%; Score 349; DB 11; Length 429;
Best Local Similarity 38.8%; Pred. No. 4.8e-27;
Matches 71; Conservative 36; Mismatches 72; Indels 4; Gaps 2;

QY 5 LMNRRKFLYQFKNVWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCHVELLFLRY 64
DB 20 LISQETEFHFKNLRYAIDRKDTFLCYEVTNRKDCSPVSLHHGVFKKNINIAEICFLYW 79

QY 65 ISDWDLPGRCYRVWTFTSWSPCYDCARHVADEFGRGNPNLSLRIFTARLYFCEDRKA 121
DB 80 FHDKVLKVLSPREEFKITWYMSWSPCFCEAEQVLRFLATHNLSLDFSSRLYNIRDPEN 139

QY 122 EPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQLRILL 181
DB 140 Q-QNLCRLVQEGAQVAAMDLYEFKKCKWKKFVDNGRRFRPWKKLLTNFRYQDSKLOEILR 198

QY 182 PLY 184
DB 199 PCY 201
```

```
RESULT 8
Q96AK3
ID Q96AK3 PRELIMINARY; PRT; 386 AA.
AC Q96AK3;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-NAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 46.6 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017022; AAH17022.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 386 AA; 46598 MW; 94C7253BDC85B22 CRC64;

Query Match 31.0%; Score 336.5; DB 4; Length 386;
Best Local Similarity 40.8%; Pred. No. 7.8e-26;
Matches 71; Conservative 27; Mismatches 71; Indels 5; Gaps 2;

QY 11 FLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRN----KNGCHVELLFLRYS 66
DB 13 FFFHKNLLKAGRNESWLCFTMEVTKHHSVFRKRGVFRNOVDPTHCHAEKFCFLSFC 272
QY 67 DWLDPGRCYRVWTFWTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPGL 126
DB 273 DDILSPNTNVEVWTSWSPCECAGEVAEFLARHSNVNLITFTARLYCFWDTQY-EGL 331
QY 127 RRLHRAGVQIAIMTFKDYFCWNTFVENHETFFKAWEGHLHNSVRLSRLRL 180
DB 332 CSLSQEGASVIMGYKDFVSCWKNFVSDDEPFKPMKGLQTNFLRLKRLREIL 385

RESULT 9
Q94V1
ID Q94V1 PRELIMINARY; PRT; 139 AA.
AC Q94V1;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NAR-2002 (TReMBLrel. 20, Last annotation update)
DE DJ494G10.1 (Novel protein similar to Phorbol 1 and APOBEC1
DE (Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1))
DE (Fragment).
GN DJ494G10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramsay H.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078641; CAB45659.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 139 AA; 16075 MW; C77CB711DDAAA9C0 CRC64;

Query Match 29.1%; Score 315.5; DB 4; Length 139;
Best Local Similarity 48.5%; Pred. No. 3.2e-24;
Matches 65; Conservative 24; Mismatches 40; Indels 5; Gaps 2;

QY 47 GYLKNGKCHVELLFLRYSWDLPGRCYRVWTFWTSWSPCYDCARHVADFLRGNPNLSL 106
DB 6 GFL---EGRHAEELFELVDPFWKLDLDQDYRVCTFTSWSPCYSCAQEMAKFISKNHVSL 62

RESULT 10
Q9WV35
ID Q9WV35 PRELIMINARY; PRT; 224 AA.
AC Q9WV35;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE APOBEC-2 protein.
GN APOBEC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=99333690; PubMed=10403781;
RA Liao W., Hong S.H., Chan B.H., Rudolph F.B., Clark S.C., Chan L.;
RT "APOBEC-2, a cardiac- and skeletal muscle-specific member of the
RT cytidine deaminase supergene family.";
RL Biochem. Biophys. Res. Commun. 260:398-404(1999).
DR EMBL; AF161699; AAD45361.1; -.
DR MGD; MGI:1343178; Apobec2.
SQ SEQUENCE 224 AA; 25660 MW; 75F98BC2CF2EBE0A CRC64;

Query Match 25.6%; Score 278; DB 11; Length 224;
Best Local Similarity 36.7%; Pred. No. 3.4e-20;
Matches 65; Conservative 26; Mismatches 74; Indels 12; Gaps 5;

QY 11 FLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKN-GCHVELLFLRYSWD 69
DB 52 FKQFRNVEYSSGRNKTFLCYVVEQSGGQAQATQCYLEDEHAGAHAEAFNTILP-A 110
QY 70 LDPGRCYRVWTFWTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEP-GL 126
DB 111 FDPALKYVNTWVWSSSPCAACADRLTKLTKNLLILVSLRFLME----EPEVQAAL 166
QY 127 RRLHRAGVQIAIMTFKDYFCWNTFVENHETFFKAWEGHLHNSVRLSRLRL 180
DB 167 KKLKAEAGCKLRIMKPDQFEYIWNQFVEQEGESKAFEPWEDIQENFLYYEEKLADIL 223

RESULT 11
Q9Y235
ID Q9Y235 PRELIMINARY; PRT; 224 AA.
AC Q9Y235;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE DJ34B21.2 (Putative novel protein similar to PART of APOBEC1
DE (PHORBOLIN 1, apolipoprotein B mRNA editing protein)).
GN DJ34B21.2 OR APOBEC-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA Peck A.;
RT "APOBEC-2, a cardiac and skeletal muscle specific member of the
RT cytidine deaminase supergene family.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
```

```
DE mRNA editing protein) and Phorbolin) (Fragment).
GN BK150C2.6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45276.1; -.
DR InterPro: IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMPC_DEAMINASES; UNKNOWN_1.
KW Lipoprotein.
FT NON_TER 1
FT NON_TER 103
SQ SEQUENCE 103 AA; 12146 MW; 5DC969AE3ED348D9 CRC64;

Query Match      23.7%; Score 257.5; DB 4; Length 103;
Best Local Similarity 50.0%; Pred. No. 1.7e-18;
Matches 49; Conservative 13; Mismatches 35; Indels 1; Gaps 1

QY 55 CHVELLFRLYSIDWLDGRCYRYTWTSNSPCYDCARHVADFLRGPNLSLRIFTARTLY 114
Db ||| |||| | | | | | | | | | | | | | | | | : : | | | | | |
7 CHAERCFLSFCDILSPNTNVEYWTWSVPCCEAGEVAEFLARHSNVNLTIPTARTLY 66
QY 115 FCEDRKAEPGLRHLRHAGVQIAIMTKDFYFCWNTFV 152
Db : | : | | | | : | : | : | | | | | | | | | | | | | | | |
67 YFWDTDYQ-BGLRSLSOEGASVEIMGYKFKYCWFNEV 103

RESULT 14
Q9UH18 ID Q9UH18 PRELIMINARY; PRT; 111 AA.
AC A90H18;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE BK150C2.1 (Putative novel Phorbolin 1 like protein) (Fragment).
GN BK150C2.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45269.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 111 AA; 13006 MW; EF8922ABFFFA7028 CRC64;

Query Match      23.1%; Score 251; DB 4; Length 111;
Best Local Similarity 46.4%; Pred. No. 8.2e-18;
Matches 51; Conservative 20; Mismatches 29; Indels 10; Gaps 3

QY 76 YRVTFWFSWSPCYD--CARHVADFLRGPNLSLRIFTARTLYFCEDRKAEPP--EGLRLH 130
Db ||||||| ||||| | | | | | | | | | | | | | | | | : | | | | |
2 YRVTFWFSWSPCFSWGCAGEVRAPLEQTHVRLIFARIY-----DYDPLYKEALQMLR 56
QY 131 RAGVQIAIMTKDFYFCWNTFVENHERTFKAWGLEHNSVRLSRQLRRIL 180
Db || : || | | : : : | | | | | | : | : | | | | | | | | | | | |
57 DAGQAQVSIMTYDEFKHCHWDITFYDHGGCPFPQPDGDLDEHSQALSGLRAIL 106

RESULT 15
Q9Y554 ID Q9Y554 PRELIMINARY; PRT; 121 AA.
AC A9Y554;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BK150C2.9 (Putative novel Phorbolin 1 like protein) (Fragment).
GN BK150C2.9.
```

GN BK150C2.9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RP [1]
 RA SEQUENCE FROM N.A.
 RA Matthews L.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL022318; CAB45272.1; -.
 FT NON_TER 1
 FT NON_TER 121 121
 SQ SEQUENCE 121 AA; 14714 MW; B8FE133AACBE6D59 CRC64;

Query Match 22.1%; Score 240.5; DB 4; Length 121;
 Best Local Similarity 39.3%; Pred. No. 1e-16; Indels 1; Gaps 1;
 Matches 48; Conservative 19; Mismatches 54; Indels 1; Gaps 1;

QY 59 LLFLRYISDWLDLPGRCYRVTVFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCED 118
 Db 1 MCFLSWFCGNRLPANRRFOITWFSWNPCLEPCVVVKTKFLAEHPNVTLTISAARLYYRD 60

QY 119 RKAPEGLRLHRRAGVQIAIMTFKDYFYCWNTFEVENHERTEKAMEGLHENSRLSRQLRR 178
 Db 61 RDWR-WVLLRLHKGARVKINDYEDFAFCWENFVCNEGQPPMPWKFDNDNYASLHRTIKE 119

QY 179 IL 180
 Db 120 IL 121

Search completed: June 14, 2003, 18:11:07
 Job time : 80 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:08:03 ; Search time 26 Seconds
(without alignments)
224.067 Million cell updates/sec

Title: US-09-966-880A-8
Perfect score: 1086
Sequence: 1 MDSLMMRRKFLYQFNVRW.....ILLPLYEVDLDAFRTLGL 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	34.0	190	1 US-08-816-241-1	Sequence 1, Appli
2	369.5	34.0	190	3 US-09-128-395-1	Sequence 1, Appli
3	271	25.0	222	1 US-08-687-895-1	Sequence 1, Appli
4	271	25.0	222	2 US-09-040-482-1	Sequence 1, Appli
5	245	22.6	116	1 US-08-687-895-3	Sequence 3, Appli
6	245	22.6	116	1 US-08-816-241-3	Sequence 3, Appli
7	245	22.6	116	2 US-09-040-482-3	Sequence 3, Appli
8	245	22.6	116	3 US-09-128-395-3	Sequence 3, Appli
9	216.5	19.9	236	1 US-08-687-895-4	Sequence 4, Appli
10	216.5	19.9	236	1 US-08-816-241-4	Sequence 4, Appli
11	216.5	19.9	236	2 US-09-040-482-4	Sequence 4, Appli
12	216.5	19.9	236	3 US-09-128-395-4	Sequence 4, Appli
13	210	19.3	236	1 US-08-158-682A-4	Sequence 4, Appli
14	198	18.2	229	1 US-08-158-682A-2	Sequence 2, Appli
15	198	18.2	229	1 US-08-015-203-2	Sequence 2, Appli
16	198	18.2	229	1 US-08-687-895-5	Sequence 5, Appli
17	198	18.2	229	1 US-08-816-241-5	Sequence 5, Appli
18	198	18.2	229	2 US-09-040-482-5	Sequence 5, Appli
19	198	18.2	229	3 US-09-128-395-5	Sequence 5, Appli
20	81	7.5	440	4 US-09-092-315-3	Sequence 3, Appli
21	80	7.4	476	4 US-09-092-315-5	Sequence 3, Appli
22	77.5	7.1	402	6 5352575-5	Patent No. 5352575
23	75.5	7.0	343	1 US-08-180-209B-56	Sequence 56, Appli
24	75.5	7.0	343	4 US-08-474-853-56	Sequence 56, Appli
25	75.5	7.0	343	4 US-09-166-205B-56	Sequence 56, Appli
26	75.5	7.0	343	5 PCT-US94-02629-56	Sequence 56, Appli
27	73.5	6.8	533	1 US-08-488-305A-6	Sequence 6, Appli

28	72.5	6.7	366	6	5470718-4	Patent No. 5470718
29	72	6.6	1094	2	US-08-680-326-40	Sequence 40, Appli
30	72	6.6	1400	1	US-08-080-255-7	Sequence 7, Appli
31	72	6.6	1400	3	US-08-465-713-7	Sequence 7, Appli
32	72	6.6	1400	5	PCT-US93-05857-7	Sequence 7, Appli
33	72	6.6	3969	4	US-08-061-376-5	Sequence 5, Appli
34	70.5	6.5	528	6	5475095-2	Patent No. 5475095
35	69.5	6.4	533	4	US-09-385-259-2	Sequence 2, Appli
36	69.5	6.4	533	4	US-09-645-370-2	Sequence 2, Appli
37	69.5	6.4	1727	2	US-08-477-451-10	Sequence 10, Appli
38	69	6.4	486	4	US-09-092-315-2	Sequence 2, Appli
39	68	6.3	425	4	US-09-092-315-6	Sequence 6, Appli
40	68	6.3	1176	1	US-07-828-788A-12	Sequence 12, Appli
41	68	6.3	1176	1	US-08-356-034-2	Sequence 2, Appli
42	68	6.3	1176	3	US-08-933-891-2	Sequence 2, Appli
43	68	6.3	1176	5	PCT-US92-11337-12	Sequence 12, Appli
44	68	6.3	1179	1	US-08-040-751-1	Sequence 1, Appli
45	67	6.2	207	4	US-09-415-522-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-816-241-1
; Sequence 1, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
; US-08-816-241-1

Query Match 34.0%; Score 369.5; DB 1; Length 190;
Best Local Similarity 44.9%; Pred. No. 1.2e-35;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

```

QY 11 FLYQFKNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
Db 17 FYQFKNLWEANDRNETWLCFTVEGIKRSVWSWKT--GVFRNQVDSETHCHAEKRCFLSW 74
QY 65 ISDWLDPGRCYRVWTFWTSWPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAPE 124
Db 75 FCDILSPNTKYQVWTSWSPCDGAGEVAEFLARHSNVNLTIFTARLYYFQ-YPCYQE 133
QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFKAWGLHENSVRLSRQLRRL 180
Db 134 GLRSLSQEGVAVEIMDYEDFKYCWENFVNDNEPFPKWLKTNFRLKRRRESL 189

RESULT 2
US-09-128-395-1
; Sequence 1, Application US/09128395
; Patent No. 6087108
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-09-128-395-1

Query Match 34.0%; Score 369.5; DB 3; Length 190;
Best Local Similarity 44.9%; Pred. No. 1.2e-35;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFKNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
Db 17 FYQFKNLWEANDRNETWLCFTVEGIKRSVWSWKT--GVFRNQVDSETHCHAEKRCFLSW 74
QY 65 ISDWLDPGRCYRVWTFWTSWPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAPE 124
Db 75 FCDILSPNTKYQVWTSWSPCDGAGEVAEFLARHSNVNLTIFTARLYYFQ-YPCYQE 133
QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFKAWGLHENSVRLSRQLRRL 180
Db 134 GLRSLSQEGVAVEIMDYEDFKYCWENFVNDNEPFPKWLKTNFRLKRRRESL 189

RESULT 3
US-08-687-895-1
; Sequence 1, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
US-08-687-895-1

Query Match 25.0%; Score 271; DB 1; Length 222;
Best Local Similarity 35.8%; Pred. No. 6e-24;
Matches 63; Conservative 25; Mismatches 76; Indels 12; Gaps 5;

QY 11 FLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKN-GCHVELLFLRYISWD 69
Db 52 FAFQFRNVEYSSGRNKTFLCYVVEAQGGQVQASRGYLEDEHAAHAEAFNTILP-A 110
QY 70 LDPGRYRVWTFWTSWPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAPE---GL 126
Db 111 FDPALRYNVTVYVSSSPCAACADRTXKTLKTNKLLILVGRLEPWE----EPDIAAL 166
QY 127 RLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWGLHENSVRLSRQLRRI 179
Db 167 KKLKAGCKLRIMKPDQFEYVQNFVEQEGESKAFQWEDIQENFLYYEKLADI 222

RESULT 4
US-09-040-482-1
; Sequence 1, Application US/09040482
; Patent No. 5916556
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME

```


Query Match 22.68; Score 245; DB 2; Length 116;

```

Query Match      22.6%; Score 245; DB 3; Length 116;
Best Local Similarity 46.4%; Pred. No. 2.8e-21;
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

QY       76 YRVTFWTSNPSCYD--CARHVADFRLGNPNLSRIETARLYFCEDRKAEP---EGLRRLLH 130
||||| ||||| : | | | : | | | | | : | | : |
Db        7 YRVTFWISNPSCFSWCAGEVRAFLGENTHVRPLPIFAARIY-----DYDPLYKEALQMLR 61
||||| ||||| : | | | : | | | | | : | | : |

QY      131 RAGVQIAIMTFFKDYFYCWNTFVENHERHTFKAMEGLHENSVRLSRQLRRIL 180
||||| ||||| : | | | | | : | | | | | : | | | |
Db       62 DAGAQVSIMTYDEFECVCTDFVYRQCQPPQPMDGLEHSHQAALSGRLLAIL 111
||||| ||||| : | | | | | : | | | | | : | | | |

RESULT 9
US-08-687-895-4
```

```
; Sequence 4, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1177798
; US-08-687-895-4

Query Match 19.9%; Score 216.5; DB 1; Length 236;
Best Local Similarity 34.9%; Pred. No. 1.7e-17;
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RDSATSFSLDFGYLR-----NKNCC-HVELLFL-RYISDWLDLPGRCYRVTFWTSWSP 86
|::: : : : | | | | : : : | : : : | : : | | | |
Db 33 RKEACLLYEIKWMSRKIRSSGKNTNHVEVNFIKFTSERDFHPSISCSITWFLSWSP 92

QY 87 CYDCARHVADFLRGNPMLSLRIFTARLYFCEDRKAEPGLRLHRAGVOIAIMTFKDYF 146
|::: : : : | : : : | : : : | : : | : : | : : | : : |
Db 93 CWECQAIREFLSRHPGVTLYVARLFWHMDQQ-NRQGLRLVNSGVTIQIMRASEYH 151

QY 147 CWNTFV 152
| | |
Db 152 CWRNFV 157

RESULT 10
US-08-816-241-4
; Sequence 4, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA

; Sequence 4, Application US/09040482
; Patent No. 5918556
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,482
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 11777906
; US-08-816-241-4

Query Match 19.9%; Score 216.5; DB 1; Length 236;
Best Local Similarity 34.9%; Pred. No. 1.7e-17;
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RDSATSFSLDFGYLR-----NKNCC-HVELLFL-RYISDWLDLPGRCYRVTFWTSWSP 86
|::: : : : | | | | : : : | : : : | : : | | | |
Db 33 RKEACLLYEIKWMSRKIRSSGKNTNHVEVNFIKFTSERDFHPSISCSITWFLSWSP 92

QY 87 CYDCARHVADFLRGNPMLSLRIFTARLYFCEDRKAEPGLRLHRAGVOIAIMTFKDYF 146
|::: : : : | : : : | : : : | : : | : : | : : | : : |
Db 93 CWECQAIREFLSRHPGVTLYVARLFWHMDQQ-NRQGLRLVNSGVTIQIMRASEYH 151

QY 147 CWNTFV 152
| | |
Db 152 CWRNFV 157

RESULT 11
US-09-040-482-4
; Sequence 4, Application US/09040482
; Patent No. 5918556
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,482
```

```

; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/687,895
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1177798
; US-09-040-482-4

```

```

Query Match 19.9%; Score 216.5; DB 2; Length 236;
Best Local Similarity 34.9%; Pred. No. 1.7e-17;
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RDSATSFSLDFGYLR-----NKNKG-HVELLFL-RYISDWLDLPGRCYRVVTWFTSWSP 86
Db 33 RKEACLLYEIKWGMRSKIRWSSGKNTTNHVEVNFIKFTSERDFHPSISCSITWFLSWSP 92

QY 87 CYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFY 146
Db 93 CWECQAIREFLSRHPGVTLLVIYVARLFWHMDQO-NRQGLRDLVNSGVTIQIMRASEYH 151

QY 147 CWNTFV 152
Db 152 CWRNFV 157

```

```

RESULT 12
US-09-128-395-4
; Sequence 4, Application US/09128395
; Patent No. 6087108
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 11777906
; US-09-128-395-4

```

```

Query Match 19.9%; Score 216.5; DB 3; Length 236;
Best Local Similarity 34.9%; Pred. No. 1.7e-17;
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RDSATSFSLDFGYLR-----NKNKG-HVELLFL-RYISDWLDLPGRCYRVVTWFTSWSP 86
Db 33 RKEACLLYEIKWGMRSKIRWSSGKNTTNHVEVNFIKFTSERDFHPSISCSITWFLSWSP 92

QY 87 CYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFY 146
Db 93 CWECQAIREFLSRHPGVTLLVIYVARLFWHMDQO-NRQGLRDLVNSGVTIQIMRASEYH 151

QY 147 CWNTFV 152
Db 152 CWRNFV 157

```

```

RESULT 13
US-08-158-682A-4
; Sequence 4, Application US/08158682A
; Patent No. 5434058
; GENERAL INFORMATION:
; APPLICANT: Davidson, Nicholas O.
; TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:
; TITLE OF INVENTION: Composition and Method
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: 321 No. 5434058th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,682A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooley, Ronald B.
; REGISTRATION NUMBER: 27,187
; REFERENCE/DOCKET NUMBER: ARCD:085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-158-682A-4

```

```

Query Match 19.3%; Score 210; DB 1; Length 236;
Best Local Similarity 35.6%; Pred. No. 9.7e-17;
Matches 48; Conservative 28; Mismatches 43; Indels 16; Gaps 5;

```

RESULT 15

Query Match	18.2%	Score 198;	DB 1;	Length 229;
Best Local Similarity	32.9%	Pred. No. 2.4e-15;		
Matches	56;	Conservative 24;	Mismatches 62;	Indels 28; Gaps 6;
Qy	24	RREYLCYVVK-----RDSATSFSLDGYLRNKGCHVELLF-L-RYISDMDLDPGR	74	
Db	33	RKECYLLEYENWGRHSIWRHTSQNT-----NK---HVEYVFTEKFTTFRYCPNT	80	
Qy	75	CYRVTWTSNCPVCDCARHVADFLGNPNLSRLFTARLYFCEDKAEPEGLRLHRAVG	134	
Db	81	RCSITWELSNSPGCECSRAITEFLSRYPHTVLFYIARLYHHADPR-NRQGLRDLISSGV	139	
Qy	135	QIALMTKDYFYCWTNVENHRTFKAWEGLHENSVRLSRQLRILLPLY	184	
Db	140	TIQMTQESQYCWNRNEVNTSPSNEAHWPYPHLWRL-----YVLELY	183	

Search completed: June 14, 2003, 18:12:28
Job time : 27 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:09:18 ; Search time 21 seconds
(without alignments)
1008.062 Million cell updates/sec

Title: US-09-966-880A-8
Perfect score: 1086
Sequence: 1 MDSLNNRRKFLYQFKNVRW.....ILLPLYEVDLDRFAFTLGL 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues
Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1086	100.0	198	9	US-09-966-880A-8
2	1008	92.8	198	9	US-09-966-880A-8
3	390	35.0	384	10	US-09-729-674-174
4	369.5	34.0	222	10	US-09-925-300-1639
5	211	19.4	229	9	US-09-966-880A-36
6	84	7.7	51	10	US-09-864-761-38853
7	81	7.5	440	9	US-10-120-319-3
8	80	7.4	476	9	US-10-120-319-5
9	80	7.4	476	10	US-09-733-524-15
10	77.5	7.1	663	9	US-10-080-960-14
11	77	7.1	790	9	US-10-153-668-164
12	75.5	7.0	382	9	US-09-847-208-25
13	75	6.9	330	9	US-10-265-593-4
14	71.5	6.6	529	9	US-10-053-192-5
15	71	6.5	440	9	US-09-910-186A-8
16	69	6.4	485	10	US-09-733-524-2
17	69	6.4	486	9	US-10-120-319-2
18	69	6.4	2015	9	US-10-066-551-1
19	68.5	6.3	295	9	US-10-042-894A-14

20	68.5	6.3	341	9	US-09-764-868-1041	Sequence 1041, Ap
21	68.5	6.3	491	10	US-09-925-302-762	Sequence 762, App
22	68	6.3	424	10	US-09-733-524-16	Sequence 16, Appl
23	68	6.3	425	9	US-10-120-319-6	Sequence 6, Appli
24	67	6.2	454	9	US-10-120-319-8	Sequence 8, Appli
25	67	6.2	479	10	US-09-733-524-18	Sequence 18, Appl
26	67	6.2	479	10	US-09-733-524-17	Sequence 17, Appl
27	67	6.2	602	10	US-09-841-132-495	Sequence 495, App
28	66.5	6.1	287	10	US-09-864-761-42836	Sequence 42836, A
29	66.5	6.1	386	10	US-09-839-497A-5	Sequence 5, Appli
30	66	6.1	440	10	US-09-733-524-3	Sequence 3, Appli
31	65.5	6.0	239	9	US-10-219-220-287	Sequence 287, App
32	65.5	6.0	1045	10	US-09-815-242-10617	Sequence 10617, A
33	65.5	6.0	4636	10	US-09-835-996A-33	Sequence 33, Appl
34	65	6.0	410	9	US-09-987-021-2	Sequence 2, Appli
35	65	6.0	410	10	US-09-957-485-2	Sequence 2, Appli
36	65	6.0	602	10	US-09-841-132-565	Sequence 565, App
37	65	6.0	637	9	US-10-041-007-24	Sequence 24, Appl
38	65	6.0	637	9	US-09-900-797-58	Sequence 58, Appl
39	65	6.0	637	10	US-09-887-586A-58	Sequence 58, Appl
40	65	6.0	637	10	US-09-903-012-58	Sequence 58, Appl
41	65	6.0	905	9	US-10-114-893-127	Sequence 127, App
42	65	6.0	1005	10	US-09-802-127-5	Sequence 5, Appli
43	64.5	5.9	250	9	US-09-880-748-1480	Sequence 1480, Ap
44	64.5	5.9	373	9	US-10-125-540-353	Sequence 353, App
45	64.5	5.9	373	9	US-10-103-313-437	Sequence 437, App

ALIGNMENTS

RESULT 1
US-09-966-880A-8
; Sequence 8, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-880A-8

Query Match 100.0%; Score 1086; DB 9; Length 198;
Best Local Similarity 100.0%; Pred. No. 3e-108;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSLNNRRKFLYQFKNVRWAKGRRETYLCYVVKRDSATSFSLDFGLNKNKCHVELL 60
Db 1 MDSLNNRRKFLYQFKNVRWAKGRRETYLCYVVKRDSATSFSLDFGLNKNKCHVELL 60
QY 61 FLRYISDWLDDPGRCYRVTFWFTSWSPCYDCARHVAADFRCGNPNLSRIFRTARLYFCSDRK 120
Db 61 FLRYISDWLDDPGRCYRVTFWFTSWSPCYDCARHVAADFRCGNPNLSRIFRTARLYFCSDRK 120
QY 121 ASPEGLRRLRHAGVQIAIMTFKDYFYCWNTFVENHRTFKAWEGHNSVRLSRQLRRL 180
Db 121 ASPEGLRRLRHAGVQIAIMTFKDYFYCWNTFVENHRTFKAWEGHNSVRLSRQLRRL 180

QY 181 LPLYEVDDLRDAFTLGL 198
Db 181 LPLYEVDDLRDAFTLGL 198

RESULT 2
US-09-966-880A-2
; Sequence 2, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966.880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-966-880A-2

Query Match 92.8%; Score 1008; DB 9; Length 198;
Best Local Similarity 92.9%; Pred. No. 6.5e-100;
Matches 183; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYYLKNKGCHVELL 60
Db 1 MDSLLMKQKFLYHFNVRWAKGRHETLYCYVVKRRDSATSFSLDFGHLRKNKSGCHVELL 60

QY 61 FLRYISDWLDPGRCYRVWFTSWSPCYDCARHVADFLRGNPNLSRIFTARLYFCEDRK 120
Db 61 FLRYISDWLDPGRCYRVWFTSWSPCYDCARHVADFLRGNPNLSRIFTARLYFCEDRK 120

QY 121 AEPEGLRRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWGLHENSRLRQLRRL 180
Db 121 AEPEGLRRLHRAGVQIGIMTFKDYFCWNTFVENHRTFKAWGLHENSRLRQLRRL 180

QY 181 LPLYEVDDLRDAFTLGL 197
Db 181 LPLYEVDDLRDAFTLGL 197

RESULT 3
US-09-729-674-174
; Sequence 174, Application US/09729674
; Patent No. US2001003935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steining II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Genetics Institute, Inc.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6055-64X
; CURRENT APPLICATION NUMBER: US/09/729.674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 174
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-729-674-174

Query Match 35.9%; Score 390; DB 10; Length 384;
Best Local Similarity 44.9%; Pred. No. 1.6e-33;
Matches 83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;

QY 6 MNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRNK-----NGC 55
Db 197 MDPPTFTFNPNPVRGRHETLYCYEVEVRHNDTWLLNQRRGFLCNAQPHKHGFLBGR 256

QY 56 HVELLFLRYISDWLDPGRCYRVWFTSWSPCYDCARHVADFLRGNPNLSRIFTARLYF 115
Db 257 HAECLFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIY- 315

QY 116 CEDRKAEPEGLRRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWGLHENSRLRQLRRL 175
Db 316 -DQGRCOEGLRLAEAGAKISIMTYSEFKHCWDTFVDHQCPQPDGDLDEHSQDLGR 374

QY 176 LRLIL 180
Db 375 LRAIL 379

RESULT 4
US-09-925-300-1639
; Sequence 1639, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Ruben
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925.300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1639
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1639

Query Match 34.0%; Score 369.5; DB 10; Length 222;
Best Local Similarity 44.9%; Pred. No. 1.3e-31;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
Db 49 FYQFNKLWEANDRNETWLCFTVEGIKRVSWSKWT--GVFRNQVDSETHCAERCFLSW 106

QY 65 ISDWLDPGRCYRVWFTSWSPCYDCARHVADFLRGNPNLSRIFTARLYFCEDRKAEPE 124
Db 107 FCDLILSPNTKYQVWTWYTSWSPCDACAGEVAEFLARHSNVLNLTFTARLYFYO-YPCYQE 165

QY 125 GLRRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWGLHENSRLRQLRRL 180
Db 166 GLRSLSQBGVAVVEIMDYEDFKYCWENFVYNDNEFPKWKGLKTNFRLKRLRESL 221

RESULT 9
 US-09-733-524-15
 ; Sequence 15, Application US/09733524
 ; Patent No. US20020068347A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Governors of the University of Alberta, a Canada Corporation
 ; APPLICANT: Taylor, Diane E.

Query Match 7.1%; Score 77.5; DB 9; Length 663;
Best Local Similarity 22.5%; Pred. No. 7.8;
Matches 42; Conservative 36; Mismatches 94; Indels 15; Gaps 6;

QY 16 KNRWAKGRREYLCVVKRRDSATSFSLDFGLRNKNGCHVELLFLRYSID---WDLP 72
Db EKIQKSGDDATLPSFLPKDKGTGTRIGDLAPQDMKKVCHLALIELTALYDLVGLQ 308
QY 73 GRCYRVTWFTSPCYDCARHVAFLRGNPNLSL-RIFTARLYFCEDRKAEPGLRLRHR 131
Db ORAVKIKTKDSGLFVPLTALLQDQKVPGMRIPLIFOKLISRIEREGLETGLRIPG 368
QY 132 AGVQIAIMTFKDYFCWNTFVENHRTFKAWGL--HENSVRSLRQLRLLPLYEVDL 189
Db AAIRIKNL-----COELEAKFYEGTFN-WESVKQHDAAALLKLFIRELPQLLSVEYL 420
QY 190 RDAFRTL 196
Db 421 K-AFAQV 426

RESULT 11
US-10-153-668-164
; Sequence 164; Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-164

Query Match 7.1%; Score 77; DB 9; Length 790;
Best Local Similarity 23.2%; Pred. No. 11;
Matches 51; Conservative 23; Mismatches 66; Indels 80; Gaps 11;

QY 29 LCYVVKRRDSATSFSL-----DFGY-----LRNKNGCHVELLFLRYSIDWDLDPGR 75
Db LCSVLKASGLQSLSDATFASPQDFGLVQLTKRYN-----LALKRLSFHDMNLAD 637
QY 76 YRVTFP-----TSWSPCYDCARHVAFL-----RGNPNL-SLRIFTARL----- 113
Db QSEVLELLONLTLQELTFSCFLFERPAQFLPEWVAAMKGNSTLGLRLPGNRLGNAGL 697
QY 114 -----YFCEDRKA-----EPEGL-----RRLHRAGVQIALMTKDFYFCWNTFV 152
Db LALADVFSEDSSSLQQLDISSNCKIPDGLLEFAKLER-----WGRGA 741
QY 153 ENHRTFKAWGLHENSVRSLRQLRLLPLYEVDL 192
Db FGHLRLFQNW--LDQDAVTAREAIRRLRATCHVVSWSWS 779

RESULT 12
US-09-847-208-25
; Sequence 25; Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: I98-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Apis mellifera (Honeybee)
US-09-847-208-25

Query Match 7.0%; Score 75.5; DB 9; Length 382;
Best Local Similarity 21.6%; Pred. No. 6.5;
Matches 36; Conservative 21; Mismatches 51; Indels 59; Gaps 7;

QY 81 FTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRKAEPGLRLHRAGVQIAIMT 140
Db 144 FESWRPIF---RQNWASLQPYKKLSVEVVRREHPFWDQVQEAKRRFEKYGQLFMEET 200
QY 141 FK-----DYFYCWN-----TFVENHRT-----FK 160
Db 201 LKAAKMRPAANWGYAYPCYNLTNPQSAQCEATTMQENDKMSLWLFESDVLPSVYL 260
QY 161 AWE-----GLHENSVRSLRQL---RRILLPL--YEVDLDR 191
Db 261 RWNLTSGERVLVGGRVKALRIARQMTTSRKKVLPVYWKYQDQRRD 307

RESULT 13
US-10-265-593-4
; Sequence 4; Application US/10265593
; Publication No. US20030049789A1
; GENERAL INFORMATION:
; APPLICANT: LADUNGA, Steven et al
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000818CON
; CURRENT APPLICATION NUMBER: US/10/265,593
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 09/708,701
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 60/243,467
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Chick
US-10-265-593-4

Query Match 6.9%; Score 75; DB 9; Length 330;
Best Local Similarity 23.6%; Pred. No. 6.1;
Matches 42; Conservative 18; Mismatches 74; Indels 44; Gaps 8;

QY 1 MDSLNRNRKFLYQFKNVRWAKORRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
Db 52 MEAISMSALECOYQFRFRMNCTLEGYRASLLKRGFKETAFL-----YATSSAG----- 101
QY 61 FLRYISDWDLDPGRCYRVT-----WFTSWSPCYDCARH-----VADFLRGNPNL 104

Db . 102 -LTHAMAKASAGRMERTCDEAPDLENREAW--QWGGCGDNLYSNKFYKFLGRKPNK 158
Qy 105 SLRIETARLYFCEDR-----KABPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHE 156
Db 159 DLR---ARVDFHNNLVGMKVIKAGVETTKCHGVSSGCTVRT-----CWRQLSPFHE 207

RESULT 14

US-10-053-192-5

; Sequence 5, Application US/10053192

; Publication No. US20030087336A1

; GENERAL INFORMATION:

; APPLICANT: BACHMANN, Heinrich

; APPLICANT: BRUGGER, Roland

; APPLICANT: FRIEDLEIN, Arno M

; APPLICANT: WIRTZ, Gabriele M

; APPLICANT: WOGGON, Wolf-Dietrich

; APPLICANT: WYSS, Adrian

; APPLICANT: WYSS, Markus

; TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID

; FILE REFERENCE: B.B-CAROTENE 15,15'-DIOXYGENASES,...

; CURRENT APPLICATION NUMBER: US/10/053.192

; CURRENT FILING DATE: 2002-01-15

; PRIOR APPLICATION NUMBER: 103382.0

; PRIOR FILING DATE: 1999-02-22

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 529

; TYPE: PRT

; ORGANISM: BOVINE

US-10-053-192-5

Query Match 6.6%; Score 71.5; DB 9; Length 529;

Best Local Similarity 20.8%; Pred. No. 26;

Matches 31; Conservative 25; Mismatches 42; Indels 51; Gaps 9;

Qy 57 VELLFLRYISDWLDPCRCYRVVTFWTSWSPCYDCAR-----HVADEFRLRGNPNLSLRI 108

Db 246 VKINLFKFLSSWSL-----WGANYMDCFEFTNETMGVWLHIADKKR----- 287

Qy 109 FTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYF-----CWN--TFVENHERTFK 160

Db 288 ---KKYLNKKYRTSPFNL--FHH-----INTYEDNGFLIVDLCCWKGFEEFYNYFTLYL 336

Qy 161 A-----WEGLHENSVRLSR-OLRRILLPL 183

Db 337 ANLRNWEVVKKARKAPQEVRRYVLPL 365

RESULT 15

US-09-910-186A-8

; Sequence 8, Application US/0910186A

; Publication No. US2003009025A1

; GENERAL INFORMATION:

; APPLICANT: U.S. Army Medical Research & Material Command

; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM

; FILE REFERENCE: A33626-A 067252.0107

; CURRENT APPLICATION NUMBER: US/09/910,186A

; CURRENT FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: PCT/US00/12890

; PRIOR FILING DATE: 2000-05-12

; PRIOR APPLICATION NUMBER: 09/611,419

; PRIOR FILING DATE: 2000-07-06

; PRIOR APPLICATION NUMBER: 60/133,865

; PRIOR FILING DATE: 1999-05-12

; PRIOR APPLICATION NUMBER: 60/133,866

; PRIOR FILING DATE: 1999-05-12

; PRIOR APPLICATION NUMBER: 60/133,867

; PRIOR FILING DATE: 1999-05-12

; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; OTHER INFORMATION: sequence
US-09-910-186A-8

Query Match 6.5%; Score 71; DB 9; Length 440;

Best Local Similarity 20.8%; Pred. No. 23;

Matches 35; Conservative 26; Mismatches 49; Indels 58; Gaps 9;

Qy 5 LMNRRKELYQFKNVWAKGRRETYL--CYVVKRRDSATSPS-----LDFGYLRN 51

Db 270 ILTRSKYNQNSKYINY----RDLYIGEKFIRRKNSQSINDDIVRKEDYIYLDFFNLNQ 325

Qy 52 KNGCHV-----ELLFLRYISDWLDPCRCYRVVTFWTSW--SPCYDCARHVADEFRLGN 101

Db 326 EWRVYKYKPKKEEKLFLAPISDS-----ELNTIOIKEYDEQPTVSC----- 370

Qy 102 PNLSLRIFTARLYFCEDRKAEP-----GLRRLHRAGVQIAIMTFKDYF 145

Db 371 -----QLLFKKDEESTDEIGLIGIHRFYESG--IVFEEVKDYF 406

Search completed: June 14, 2003, 18:12:57

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 19, 2003, 11:55:51 ; Search time 2485 Seconds
(without alignments)
2318.855 Million cell updates/sec

Title: us-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLLMNRKFLQYKVRW.....ILLPLYEVDDLDRDAFRTLGL 198

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09966880/runat_14062003_175645_16271/app_query.fasta_1.391
-DB=Genembl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09966880.@CGN_1.1_2496.erunat_14062003_175645_16271 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.in:*
18: em.hum:*
19: em.mu:*
20: em.Or:*
21: em.Om:*
22: em.Ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rod:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1086	100.0	596	9	AF529823 Homo sapi
2	1086	100.0	596	9	AF529826 Homo sapi
3	1086	100.0	596	12	AF529837 Mus muscu
4	1086	100.0	1837	9	BC006296 Homo sapi
5	1086	100.0	2791	9	AB040431 Homo sapi
6	1084	99.8	596	9	AF529815 Homo sapi
7	1083	99.7	596	12	AF529833 Mus muscu
8	1083	99.7	596	12	AF529835 Mus muscu
9	1082	99.6	596	12	AF529830 Mus muscu
10	1082	99.6	596	12	AF529831 Mus muscu
11	1082	99.6	596	12	AF529838 Mus muscu
12	1081	99.5	596	9	AF529819 Homo sapi
13	1081	99.5	596	9	AF529822 Homo sapi
14	1081	99.5	596	9	AF529825 Homo sapi
15	1081	99.5	596	12	AF529829 Mus muscu
16	1081	99.5	596	12	AF529832 Mus muscu
17	1080	99.4	595	12	AF529834 Mus muscu
18	1080	99.4	596	9	AF529827 Homo sapi
19	1079	99.4	596	9	AF529820 Homo sapi
20	1079	99.4	596	9	AF529821 Homo sapi
21	1079	99.4	597	12	AF529828 Mus muscu
22	1078	99.3	596	12	AF529836 Mus muscu
23	1078	99.3	596	12	AF529839 Mus muscu
24	1076	99.1	597	12	AF529842 Cricetulu
25	1076	99.1	597	12	AF529843 Cricetulu
26	1076	99.1	597	12	AF529844 Cricetulu
27	1076	99.1	597	12	AF529847 Cricetulu
28	1076	99.1	597	12	AF529852 Cricetulu
29	1075	99.0	596	9	AF529824 Homo sapi
30	1072	98.7	596	9	AF529816 Homo sapi
31	1072	98.7	596	9	AF529818 Homo sapi
32	1072	98.7	597	12	AF529845 Cricetulu
33	1071	98.6	597	12	AF529851 Cricetulu
34	1070	98.5	596	9	AF529817 Homo sapi
35	1070	98.5	597	12	AF529841 Cricetulu
36	1070	98.5	597	12	AF529846 Cricetulu
37	1069	98.4	597	12	AF529848 Cricetulu
38	1069	98.4	597	12	AF529849 Cricetulu
39	1067	98.3	597	12	AF529850 Cricetulu
40	1065	98.1	597	12	AF529853 Cricetulu
41	1065	98.1	597	12	AF529854 Cricetulu
42	1057	97.3	591	12	AF529855 Cricetulu
43	1044	96.1	577	12	AF529840 Mus muscu
44	1008	92.8	2440	10	AF132979 Mus muscu
45	986	90.8	547	12	AF529856 Cricetulu

ALIGNMENTS

```
AF529823      596 bp      mRNA      linear      PRI 19-AUG-2002
LOCUS      Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.
DEFINITION      AF529823
ACCESSION      AF529823
VERSION      AF529823.1 GI:22297233
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 596)
AUTHORS      Martin.A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
JOURNAL      cells
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin.A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source      1..596
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="Ramos 9"
              /gene="AID"
              /product="AID"
              /codon_start=1
              /protein_id="AA095410.1"
              /db_xref="GI:22297234"
              /translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
              DFGYLRNKGCHVELLYRISDWLDPRGCRVYTWFTSWSPCYCARHVADEFRLGNP
              NLSLRIFARLYFCEDRAPEGLRHLRHAGVQIAIMTFKDYFCWNTFVENHETFK
              AWEGLHENSVRSLRQLRILLPLYEVDLDRDAFTLGL"
BASE COUNT      128 a 163 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.:      3.81e-109      Length:      596
Score:      1086.00      Matches:      198
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      9      Gaps:      0

US-09-966-880A-8 (1-198) x AF529823 (1-596)

QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACAGCCTATTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60

QY      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACA 120

QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCAGTGGAATGCTC 180

QY      61 PheLeuArgTyrTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      181 TTCCTCCGCTACATCTCGACTGGGACCTAGACCCCTGCCGCTGCTACCGCGTACCCTGG 240

QY      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      241 TTCACCTCTGGAGCCCTGTGTACACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300

QY      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      1 ATGGACAGCCTCTTATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
```

```
Db      301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACGCGAAG 360
QY      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      361 GCTGAGCCCGAGGGCTGCGCGCGCTGCACCGCGGGGTGCAAAATAGCATCATGACC 420
QY      141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      421 TTCAAAGATTAATTTTACTGCTGGAATACTTTGTAGAAAACCATCAAGAAGACTTTCAAA 480
QY      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180
Db      481 GCGTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540

QY      181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      541 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGTAATTTGGGACTT 594

RESULT 2
AF529826      596 bp      mRNA      linear      PRI 19-AUG-2002
LOCUS      Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds.
DEFINITION      AF529826
ACCESSION      AF529826
VERSION      AF529826.1 GI:22297239
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 596)
AUTHORS      Martin.A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
JOURNAL      cells
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin.A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source      1..596
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="Ramos 12"
              /gene="AID"
              /product="AID"
              /codon_start=1
              /protein_id="AA095413.1"
              /db_xref="GI:22297240"
              /translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
              DFGYLRNKGCHVELLYRISDWLDPRGCRVYTWFTSWSPCYCARHVADEFRLGNP
              NLSLRIFARLYFCEDRAPEGLRHLRHAGVQIAIMTFKDYFCWNTFVENHETFK
              AWEGLHENSVRSLRQLRILLPLYEVDLDRDAFTLGL"
BASE COUNT      127 a 163 c 155 g 151 t
ORIGIN
Alignment Scores:
Pred. No.:      3.81e-109      Length:      596
Score:      1086.00      Matches:      198
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      9      Gaps:      0

US-09-966-880A-8 (1-198) x AF529826 (1-596)

QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACAGCCTCTTATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60
```

```

QY 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAGGGCTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGTGACAGTGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCACAGTGGAAATGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTTCCCTCATCTCGGACTGGGACCTAGACCTGGCGGTGCTACCGGTCACTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTTGGGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AATCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGGCGGTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCAAGAACTTTCAAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 540
QY 181 LeuProLeuTyrGluValAspLeuAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTT 594

RESULT 3
AF529837
LOCUS
DEFINITION Mus musculus clone 10 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529837
VERSION AF529837.1 GI:22297261
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="10"
/cell_type="hybridoma Pl-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1

source
BC006296
1837 bp mRNA linear PRI 12-JUL-2001
Homo sapiens, activation-induced cytidine deaminase, clone
MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION BC006296
VERSION BC006296.1 GI:13623400
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1837)

```

```

/product="AID"
/protein_id="AA095424.1"
/db_xref="GI:22297262"
/translation="MDSLMMRRKFLYQFKVNRWAKGRRETYLCYVVRDSDATSFSL
NFGYLNRNKGCHVELLFYRISDWDLPGRCYRVTFWFTSMSPCYDCARRHVAOFLRGNP
DLRSRIETFARLYEDRKAEPGRRLRHAGVQIAIMTFKDYFYCWNTFVENHETFK
AWEGIHENSVRLSQRLRILLPLLYEVDLDRAFTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Alignment Scores:
Score: 3.81e-109 Length: 596
Pred. No.: 1086.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529837 (1-596)

```

```

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACACCTCTTGTATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCCGCTGG 60
QY 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAGGGCTGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCACAGTGTCTACA 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTTCCCTCATCTCGGACTGGGACCTAGACCTGGCGGTGCTACCGGTCACTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTTGGGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGGCTGGCGGTGCACCGCGGGGTGCAAAATAGCCATCATGACT 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCAAGAACTTTCAAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 540
QY 181 LeuProLeuTyrGluValAspLeuAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTT 594

```

```

RESULT 4
BC006296
LOCUS
DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone
MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION BC006296
VERSION BC006296.1 GI:13623400
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1837)

```

AUTHORS TITLE JOURNAL

Strausberg, R.
Direct Submission
Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 17 Row: a Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9988409.

FEATURES

source

1. .1837
/organism="Homo sapiens"
/db_xref="LocusID:57379"
/db_xref="taxon:9606"
/clone="MGC:12911 IMAGE:4054915"
/tissue_type="Primary B-Cells from Tonsils"
/clone_lib="NIH_MGC_48"
/lab_host="DH10B-R"
/notes="Vector: pOT87"
86. .582
/codon_start=1
/product="activation-induced cytidine deaminase"
/protein_id="AAH06296.1"
/db_xref="GI:13623401"
/translation="MDSLMMNRKELYQFNVRWAKRRETYLVCYVVKRDSATSFSL
DFGLRNKNGCHVELLEFLYISDNDLDPCRCYVWFTWSPCYDCARHVAFLRNP
NLSRLFTARLYFCEDRKAEPEGLRHLRAGVOIAIMTFKDYFCWNTFVENHRTFK
AWEGHENSVRLSRQLRIILLPLYEVDLDRDAFTLGL"

CDS

BASE COUNT 530 a 387 c 421 g 499 t
ORIGIN

Alignment Scores:

Pred. No.: 1.55e-108 Length: 1837
Score: 1086.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-966-880a-8 (1-198) x BC006296 (1-1837)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyTGlnPheLysAsnValArgTtrp 20
|||||
DB 86 ATGCACAGCTCTGTATGACCGGAGAGTTCTTTACCAATTCAAAATGTCCGCTGG 145
QY 21 AlaLysGlyArgArgGluThrTyLeuCyTyrValValLysArgArgAspSerAlaThr 40
|||||
DB 146 GCTAAGGGTGGCGGTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACGTGTACA 205
QY 41 SerPheSerLeuAspPheGlyTyLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|||||
DB 206 TCCTTTTCTACTGGACTTGTGGTTATCTTCGCAATAAGAACGGCTGCCACGTTGGAATTGCTC 265

QY 61 PheLeuArgTyTrrIleSerAspTrpAspLeuAspProGlyArgCysTyArgValThrTrp 80
|||||
DB 266 TTCCTCCGCTACATCTCGGACTGGACCTAGACCCTGGCGCTGTACCGCTCAGCTGG 325
QY 81 PheThrSerTrpSerProCysTyArgCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
DB 326 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCCAGGG 385
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyPheCysGluAspArgLys 120
|||||
DB 386 AACCCCAACCTCAGTCTGAGGACTTTCACCGCGCGCTTCTTCTGTGAGGACCCCAAG 445
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
DB 446 GCTGACCCCGAGGGCTGCGGGGCTGCACCGCGCGGGGTGCAATATGCCATCATGACC 505
QY 141 PheLysAspTyPheTyTrpCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
DB 506 TTCAAAGATTATTTTACTGCTGGATACTTTTGTAGAAACCATGAAAGAACTTTTCAA 565
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGlyLeu 180
|||||
DB 566 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCCATCCCTT 625
QY 181 LeuProLeuTyTrrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
|||||
DB 626 TTGCCCTCTATGAGGTTCATGACTTACGAGAGCATTTTCGTACTTTGGGACTT 679
RESULT 5
LOCUS AB040431 2791 bp mRNA linear PRI 03-OCT-2000
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,
complete CDS.
ACCESSION AB040431
VERSION AB040431.1 GI:9988409
KEYWORDS AID; activation-induced cytidine deaminase; Human AID.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.
TITLE Isolation, tissue distribution, and chromosomal localization of the
human activation-induced cytidine deaminase (AID) gene
JOURNAL Genomics 68 (1), 85-88 (2000)
MEDLINE 20408890
REFERENCE 2 (sites)
AUTHORS Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O.,
Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,
Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N.,
Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.
and Durandy, A.
TITLE Activation-induced cytidine deaminase (AID) deficiency causes the
autosomal recessive form of the Hyper-IgM syndrome (HIGM2)
JOURNAL Cell 102 (5), 565-575 (2000)
MEDLINE 20460541
REFERENCE 3 (bases 1 to 2791)
AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.
TITLE Direct Submission
Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department
of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,
Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp,
Tel: 81-75-753-4371 (ex. 4371), Fax: 81-75-753-4388)
FEATURES
source 1. .2791
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1. .2791
/gene="AID"
CDS 77..673
/gene="AID"
/codon_start=1


```

/product="activation-induced cytidine deaminase"
/protein_id="BAB1271.1"
/db_xref="GI:9988410"
/translation="MDSLNNRRKFLYQFKNVRWAKGRRETLYCYVVKRRDSATSPSL
DFGYLRNKGCHVELLFLRYISDWLDPCGRVRYTWTFSWSPCYDCARHVAQDFLRNG
NLSLRITFARLYFCEDRKAPCEPLRHAGVQIAIWTFKDYFYCWNTFVENHERIFPK
ANEGLHENSRLSRQLRRILLPLYEVDDLDRAFTLGL"
BASE COUNT      842 a  348 c  625 g  776 t
ORIGIN

Alignment Scores:
Pred. No.:      2 61e-108      Length:      2791
Score:          1086.00      Matches:      198
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              9      Gaps:      0

US-09-966-880A-8 (1-198) x AB040431 (1-2791)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
   |||||
Db 77 ATGGACAGCCTCTTGATGAACCGGAGGAGATTCTTTACCAATTCAAAAATGTCGCTGG 136

QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTysrValValLysArgArgSerAlaThr 40
   |||||
Db 137 GCTAAGGGTCGGCGTGAGACCTACCTGCTGTGTAGCTAGTGAAGAGGGGTGACAGTGTCTACA 196

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
   |||||
Db 197 TCCTTTTCACTGACATTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATTTGCTC 256

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
   |||||
Db 257 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 316

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
   |||||
Db 317 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCCGACACTGTGCCGCGACTTTCTTCGAGGG 376

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
   |||||
Db 377 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTCTGTGAGGACCCGAAG 436

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
   |||||
Db 437 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 496

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
   |||||
Db 497 TTCAAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 556

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgAlaIleLeu 180
   |||||
Db 557 GCCTGGAGAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCCTT 616

QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
   |||||
Db 617 TTGCCCCCTGTATGAGGTTGTATGACTTACGAGAGCGCATTCGTACTTTGGGACTT 670

RESULT 6
AF529815
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Martin, A. and Scharff, M.D.

```

.TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 596)

AUTHORS Martin,A. and Scharff,M.D.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES

source

1..596

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Ramos 1"

1..>596

/gene="AID"

1..>596

/gene="AID"

/note="integrated into Burkitt's lymphoma cell line Ramos"

/codon_start=1

/product="AID"

/protein_id="AA095402.1"

/db_xref="GI:2297218"

/translation="WDSLNNRRFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLEFRYLSNDMDLDFGRCYRTWFTSWSPCYDCARHVAFLRGNP
NLSLRTARLYCFCEDRKAPGELRLRHAGVQIAIMTFKDYFYCWNTFVENHERTEFK
AMEGLHENSVLRSQIRILLPLYEVDLRDAFRTGL"

BASE COUNT 128 a 163 c 155 g 150 t

ORIGIN

Alignment Scores:

Pred. No.: 6,3e-109 Length: 596

Score: 1084.00 Matches: 197

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 99.49% Mismatches: 0

Query Match: 99.82% Indels: 0

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529815 (1-596)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20

Db 1 ATGGACAGCCTCTTGATGACACGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 60

QY 21 AlaLysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr 40

Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACAGTGTCTACA 120

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

Db 121 TCGTTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGTGGAATGTCTC 180

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrrp 80

Db 181 TTCCCTCCGTACATCTCGACTGGGACCTAGACCTGGCCGCTGCTACCGCGTCCACTGG 240

QY 81 PheThrSerTrpSerProCystTyrAspCysAlaArgHisValAlaAspPheLeuArggly 100

Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120

Db 301 AAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTrrp 140

Db 361 GCTGAGCCCGAGGGGCTCGCGCGTGCACCGCGCGGGTGC AAAATAGCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCystTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160

Db 421 TTCAAGATTATTTTACTGCTGGAAATCTTTTGTAGAAAACCATGAAGAAGCACTTTCAA 480

QY 161 AlaTrrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180

ORIGIN	FEATURES	Location/Qualifiers
Alignment Scores: Pred. No.: Score: Percent Similarity: Best Local Similarity: Query Match: DB:	source 1. .596 /organism="Mus musculus" /db_xref="taxon:10090" /clone="3" /cell_type="hybridoma pl-5" /transgenic 1. .596 /organism="Homo sapiens" /db_xref="taxon:9606" 1. .>596 /gene="AID" 1. .>596 /gene="AID" /codon_start=1 /product="AID" /protein_id="AA095417.1" /db_xref="GI:22297248" BASE COUNT ORIGIN	1. .596 /organism="Mus musculus" /db_xref="taxon:10090" /clone="3" /cell_type="hybridoma pl-5" /transgenic 1. .596 /organism="Homo sapiens" /db_xref="taxon:9606" 1. .>596 /gene="AID" 1. .>596 /gene="AID" /codon_start=1 /product="AID" /protein_id="AA095417.1" /db_xref="GI:22297248" 126 a 164 c 155 g 151 t
US-09-966-880A-8 (1-198) x AF529835 (1-596)		
QY 1	MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20	
Db 1	ATGGACAGCCTCTGATGAACCGGAGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60	
QY 21	AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40	
Db 61	GCTAAGGTCGGCTGAGACCTACCTAGTGTAGTGAAGAGCGTGACATGCTACA 120	
QY 41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60	
Db 121	TCCTTTTCACTGACCTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATGCTC 180	
QY 61	PheLeuArgTyrIleSerAspTrrpAspLeuAspProGlyArgCysTyrArgValThrTrrp 80	
Db 181	TTCTCTCCGTACATCTCGGACCTGGGACCTAGACCCCTGGCGCGTGTCTACCGCTCACCTGG 240	
QY 81	PheThrSerTrrpSerProCysTyrAspCysAlaAarqHisAlaAaspPheLeuArgGly 100	
Db 241	TTACACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACCTTCTTGCAGGG 300	
QY 101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120	
Db 301	AAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360	
QY 121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValIcInIleAlaIleMetThr 140	
Db 361	GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCATAATAGCATCATGTCC 420	
QY 141	PheLysAspTyrPheTyrCysTrrpAsnThrPheValGluAsnHisGluArgThrPheLys 160	
Db 421	TTCAAGATATATTTTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTTCAA 480	
QY 161	AlaTrrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180	
Db 481	GCCTGGGAAGGGCTGCATGAAAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCCTT 540	
QY 181	LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198	
RESULT 9		
AF529830		
LOCUS	596 bp mRNA linear SYN 19-AUG-2002	
DEFINITION	Mus musculus clone 3 transgenic Homo sapiens AID (AID) mRNA, complete cds.	
ACCESSION	AF529830	
VERSION	AF529830.1	
KEYWORDS	GI:22297247	
SOURCE		
ORGANISM	house mouse.	
REFERENCE		
AUTHORS	Mus musculus	
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
JOURNAL	1 (bases 1 to 596)	
REFERENCE	Martin, A. and Scharff, M.D.	
AUTHORS	Somatic hypermutation of the AID transgene in B and non-B cells	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 596)	
REFERENCE	Martin, A. and Scharff, M.D.	
AUTHORS	Direct Submission	
TITLE		
JOURNAL		

Db	541	TTGCCCTCTATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTT	594
RESULT 10			
AF529831			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
source			
gene			
CDS			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-09-966-880A-8 (1-198) x AF529831 (1-596)			
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTirp	20
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTACCAATCAAAAATGTCGCGTGG	60
QY	21	AlaLysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAAGGGTCGCGGTGAGACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCACTGGACTTTGGTTAICTTCGCAATAGAACGGCTGCCAGTGGAAATTGCTC	180
QY	61	PheLeuArgTyrIleSerAspTirpAspLeuAspProGlyArgCysTyrArgValThrTirp	80
Db	181	TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCTGCGCCCTGCTACCGCGTCACTGG	240
QY	81	PheThrSerTirpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly	100
Db	241	TTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTTCTCGGAGGG	300
QY	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys	120
Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360
QY	121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr	140
Db	361	GCTGAGCCCGAGGGGCTGCCGCGCTGCACCGCCCGGGTGCAATAGGCATCATGACC	420
QY	141	PhelysAspTyrPheTyrCysTirpAsnThrPheValGluAsnHisGluArgThrPheLys	160
Db	421	TTCAAGATATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAACACTTTTCAA	480
QY	161	AlaTirpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu	180
Db	481	GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGTTTCGGCGCATCCTT	540
QY	181	LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu	198
Db	541	TTGCCCTCTATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTT	594
RESULT 11			
AF529838			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
source			
gene			
CDS			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-09-966-880A-8 (1-198) x AF529831 (1-596)			
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTirp	20
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTACCAATCAAAAATGTCGCGTGG	60
QY	21	AlaLysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAAGGGTCGCGGTGAGACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCACTGGACTTTGGTTAICTTCGCAATAGAACGGCTGCCAGTGGAAATTGCTC	180
QY	61	PheLeuArgTyrIleSerAspTirpAspLeuAspProGlyArgCysTyrArgValThrTirp	80
Db	181	TTCTCCCGCTACATCTCGGACTGGGACCTAGACCCTGCGCCCTGCTACCGCGTCACTGG	240

Score: 1082.00 Matches: 197
 Percent Similarity: 99.49% Conservative: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 99.63% Indels: 0
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529838 (1-596)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
 |||||
 Db 1 ATGGACACCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
 |||||
 Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 |||||
 Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGCTAGTGTAGTGAAGAGCGGTGACAGTCTACA 120
 |||||
 Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 |||||
 Db 121 TCCTTTTCACCTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGCTGGAATTTGCTC 180
 |||||
 Qy 61 PheLeuArgTyrIleSerAspTrrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 |||||
 Db 181 TTCCTCCGCTACATCTCGACTGGACCTAGACCTGGCCGCTGCTACCGCGTCACTCTGG 240
 |||||
 Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 |||||
 Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTTCGGAGGG 300
 |||||
 Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 |||||
 Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGTGAGGACCGCAAG 360
 |||||
 Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 |||||
 Db 361 GCTAGCCCGAGGGCTGCGCGCTGACCGCGCGGGGTACAAATAGTCATCATGACC 420
 |||||
 Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 |||||
 Db 421 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAAA 480
 |||||
 Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 |||||
 Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCTCTT 540
 |||||
 Qy 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 |||||
 Db 541 TTGCCCCGTATGAGGTGATGACTTACGACAGCGCATTTCTGACTTTTGGGACTT 594
 |||||

RESULT 12
 AF529819 Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.
 LOCUS AF529819.1 GI:22297225
 DEFINITION Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.
 ACCESSION AF529819.1
 VERSION AF529819.1
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 596)
 Martin, A. and Scharff, M.D.
 Somatic hypermutation of the AID transgene in B cells and non-B
 cells
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 596)
 Martin, A. and Scharff, M.D.
 Direct Submission
 TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
 Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
 JOURNAL Location/Qualifiers
 FEATURES
 Source 1..596
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

gene /clone="Ramos 5"
 CDS 1..>596
 /gene="AID"
 1..>596
 /gene="AID"
 /note="Integrated into Burkitt's lymphoma cell line Ramos"
 /codon_start=1
 /product="AID"
 /protein_id="AA095406.1"
 /db_xref="GI:22297226"
 /translation="MDSLMMRRKFLYQFKNVRWAKGRRETYLCIVVRRDSATSFSL
 DFGYLNRNKGCHVELLFYISDWLDPRGCRVTFWTSWSPCYCARHVAFLRGNP
 NLSLRIFTARLYFCEDHKAEPGLRHLRAGVQIAIMTFKDYCWNTFVENHERTFK
 AWEGLHENSRLSROLRILLPLYEVDLDRFAFTLGL"
 BASE COUNT 128 a 164 c 154 g 150 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.34e-108 Length: 596
 Score: 1081.00 Matches: 197
 Percent Similarity: 99.49% Conservative: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 99.54% Indels: 0
 DB: 9 Gaps: 0
 US-09-966-880A-8 (1-198) x AF529819 (1-596)
 Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
 |||||
 Db 1 ATGGACACCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
 |||||
 Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 |||||
 Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTCTACA 120
 |||||
 Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 |||||
 Db 121 TCCTTTTCACCTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGCTGGAATTTGCTC 180
 |||||
 Qy 61 PheLeuArgTyrIleSerAspTrrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 |||||
 Db 181 TTCCTCCGCTACATCTCGACTGGACCTAGACCTGGCCGCTGCTACCGCGTCACTCTGG 240
 |||||
 Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 |||||
 Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTTCGGAGGG 300
 |||||
 Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 |||||
 Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGTGAGGACCGCAAG 360
 |||||
 Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 |||||
 Db 361 GCTAGCCCGAGGGCTGCGCGCTGACCGCGCGGGGTACAAATAGTCATCATGACC 420
 |||||
 Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 |||||
 Db 421 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAAA 480
 |||||
 Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 |||||
 Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCTCTT 540
 |||||
 Qy 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 |||||
 Db 541 TTGCCCCGTATGAGGTGATGACTTACGACAGCGCATTTCTGACTTTTGGGACTT 594
 |||||
 RESULT 13
 AF529822 Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.
 LOCUS AF529822
 DEFINITION Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.
 ACCESSION AF529822
 VERSION AF529822.1 GI:22297231

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 596)
JOURNAL Martin,A. and Scharff,M.D.
REFERENCE Somatic hypermutation of the AID transgene in B cells and non-B
AUTHORS cells
TITLE Unpublished
JOURNAL 2 (bases 1 to 596)
REFERENCE Martin,A. and Scharff,M.D.
AUTHORS Direct Submission
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 8"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95409.1"
/db_xref="GI:22297232"
/translation="MDSLMMNRKFLYQFNVRWAKGRREYLCYVVKRRDRATSFSL
DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGLUHENSVRSLRQLRLLPLLYEVDLDRAPFTLGL"
BASE COUNT 126 a 165 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 1.34e-108 Length: 596
Score: 1081.00 Matches: 197
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.54% Indels: 0
DB: 9 Gaps: 0
US-09-966-880A-8 (1-198) x AF529822 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACCTGCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTGGTTATCTTCGCATAGAAGCGGTGCCACGTGGAATGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCCTCCGCTACATCTCGACTGGACCTAGACCTGCGCGCTGCTACCGCTCACCTGG 240
QY 81 PheThrSerTyrSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGAGGCCCTCTGCTACGACGTGCGCCGACATGTGGCGACTTTCTGCGAGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTCAGGATCTTCACCGCGCGCCCTCTACTTCTGTGAGACCGCAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTCGCGCGGTGACCGCGCGGGGTGCAAAATGACCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGCGTGCATGAAATTCAGTTCGTCTCTCCACAGACGCTTCGGCGCATCCT 540
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGTGTGATGACTTACGAGACGCACTTTCGTACTTTGGGACTT 594
RESULT 14
AF529825
LOCUS Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
DEFINITION AF529825
ACCESSION AF529825
VERSION AF529825.1 GI:22297237
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B cells and non-B
REFERENCE cells
AUTHORS Unpublished
TITLE 2 (bases 1 to 596)
JOURNAL Martin,A. and Scharff,M.D.
REFERENCE Direct Submission
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 11"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95412.1"
/db_xref="GI:22297238"
/translation="MDSLMMNRKFLYQFNVRWAKGRREAYLCYVVKRRDRSATSFL
DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGLUHENSVRSLRQLRLLPLLYEVDLDRAPFTLGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 1.34e-108 Length: 596
Score: 1081.00 Matches: 197
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.54% Indels: 0
DB: 9 Gaps: 0
US-09-966-880A-8 (1-198) x AF529825 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACAGTGTCTACA 120

```
QY 41 SerPheSerLeuAspPheLeuTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCCTTTTACAGCTGGCTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGCGGCGCTGCACCGCGGGGTGCAATAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGGCATCCTT 540
QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTT 594

RESULT 15
AF529829 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 2 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529829
VERSION AF529829.1 GI:22297245
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 596)
JOURNAL Martin, A. and Scharff, M.D.
REFERENCE Somatic hypermutation of the AID transgene in B and non-B cells
AUTHORS 2 (bases 1 to 596)
JOURNAL Unpublished
AUTHORS Martin, A. and Scharff, M.D.
REFERENCE Direct Submission
AUTHORS Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="2"
/cell_type="hybridoma Pl-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/product="AID"
/protein_id="AA095416.1"
/db_xref="GI:22297246"
/translation="MDRLIMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFYIISDWLDIPGRYRVTFWTSWSPCYDCARHVAQFLRGNP
```

```
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Alignment Scores: 1.34e-108 Length: 596
Pred. No.: 1081.00 Matches: 197
Score: 1081.00
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.54% Indels: 0
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529829 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrClnPheLysAsnValArgTrp 20
Db 1 ATGGAGCAGACTCTTGTATGAACCGGAGAAAGTTCTTTTACCAATTCAAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGCTCGGCTGAGACCTACCTGTGCTAGTAGTAGAAGAGCGTGACAGTGCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGCGGCGCTGCACCGCGGGGTGCAATAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGGCATCCTT 540
QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTT 594
```

Search completed: June 19, 2003, 18:39:41
Job time : 2488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 19, 2003, 11:52:06 ; Search time 227 Seconds
(without alignments)
1964.298 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLMMRRKFLYQFKNVRW.....ILLPLYVDLDRFRTGL 198

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US0966880/runat_14062003_175645_16263/app_query.fasta_1.391
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -IOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -NAXLEN=2000000000
-USER=US0966880 @cgn_1.1.263 @runat_14062003_175645_16263 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002:*
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1086	100.0	2818	21	AAC55312	Mouse activation-1
2	1008	92.8	2440	21	AAG55307	Human activation-1
3	644.5	59.3	6564	21	AAC55314	Human activation-1
4	644.5	59.3	11204	21	AAC55339	Human activation-1
5	505	46.5	271	21	AAC55317	Human activation-1
6	390	35.9	1534	20	AAZ20856	Polynucleotide seq
7	390	35.9	1534	22	AA59293	Human cDNA encodin
8	390	35.9	1534	24	AA90962	Human polynucleoti
9	388	35.7	1446	21	AAA12409	cDNA encoding a hu
10	369.5	34.0	610	19	AAVA8231	Human RNA editing
11	369.5	34.0	610	21	AAAT2058	cDNA encoding huma
12	369.5	34.0	950	22	AA541420	cDNA encoding nove
13	369.5	34.0	987	21	AAF16264	Human prostate can
14	369.5	34.0	987	22	AA341407	cDNA encoding nove
15	369.5	34.0	1120	22	AA159847	Human polynucleoti
16	369.5	34.0	1143	22	AA158061	Human polynucleoti
17	363.5	33.5	1348	24	ABN96785	Gene #3283 used to
18	349.5	32.2	944	24	ABL99876	Human secretory po
19	336.5	31.0	1055	24	AAD24392	Human RNA metaboli
20	322.5	29.7	716	20	AA215848	Human gene express
21	321.5	29.6	819	22	AA541669	cDNA encoding nove
22	308.5	28.4	2151	24	AA562572	cDNA sequence #359
23	277.5	25.6	1164	24	AAD27214	Human AAD45360 pro
24	274	25.2	675	24	AAD25771	Human APOBEC2 cDNA
25	274	25.2	4812	22	AAK77995	Human immune/haema
26	274	25.2	4812	22	AAK84817	Human immune/haema
27	274	25.2	12600	24	AAD25770	Human APOBEC2 gene
28	271	25.0	892	19	AAV17184	cDNA encoding a no
29	270	24.9	12600	24	AAD25817	Human APOBEC2 gene
30	262	24.1	6131	22	AA342187	Genomic sequence #
31	262	24.1	6131	22	AAK69784	Human immune/haema
32	262	24.1	201143	24	ABK83568	Human DNA differen
33	260	23.9	148	21	AAC55316	Human activation-1
34	259.5	23.9	1567	22	AA193080	Human polynucleoti
35	212	19.5	116	21	AAC55318	Human activation-1
36	209	19.2	2338	22	AAH17654	Human cDNA sequenc
37	203	18.7	454	21	AAC00672	Human secreted pro
38	199	18.3	879	15	AAQ71633	Apo-B RNA editing
39	193	17.8	879	15	AAQ71632	Apo-B RNA editing
40	192	17.7	674	24	AA562024	Porcine muscular s
41	187	17.2	300	20	AA213088	Human gene express
42	182.5	16.8	650	24	ABQ60647	Human colon cancer
43	167	15.4	445	22	AAC91325	Human polynucleoti
44	162.5	15.0	572	22	ABA63759	Human foetal liver
45	162.5	15.0	572	22	ABA30950	Probe #9416 for ge

ALIGNMENTS

RESULT 1

AAC55312
ID AAC55312 standard; CDNA; 2818 BP.

XX AAC55312;

XX 05-FEB-2001 (first entry)

DT Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.

DE Activation-induced cytidine deaminase; AID; cytidine deaminase;

KW Immune related disease; allergy; allergic disease; anti-allergic;

KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatologic;

KW gene therapy; B cell associated immune system disorder; food allergy;

KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;

KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;

KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;

KW ataxia telangiectasia; common variable immunodeficiency disorder;

KW major histocompatibility class II deficiency disease;
 XX auto immunodeficiency syndrome; IgG subclass selection disorder; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 XX CDS 80..676

FT /*tag= a
 FT /product= "activation-induced cytidine deaminase"

XX WO200058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

XX 24-JUN-1999; 99JP-0178999.

XX 27-DEC-1999; 99JP-0371382.

XX (NIBS) JAPAN TOBACCO INC.

XX (HONJ/) HONJO T.

XX Honjo T, Muramatsu M;

XX WPI; 2000-611715/58.
 XX P-PSDB; AAB24198.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 XX a target for drug development for immune-related diseases including
 XX allergies -

XX Claim 3; Page 135-139; 174pp; Japanese.

XX The present sequence encodes human activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, Digeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class)
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.

XX SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;

XX Alignment Scores:

Pred. No.: 2,74e-122 Length: 2818
 Score: 1086.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-966-880A-8 (1-198) x AAC55312 (1-2818)

QY 1 MetAspSerLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
 Db 80 ATGGACAGCCTTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGGG 139
 QY 21 AlaLysGlyArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
 Db 140 GCTAGGGTGGCGTGACCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTCTACA 199
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
 XX

Db 200 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGAATGCTC 259
 QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 Db 260 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGTCACCTGG 319
 QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 Db 320 TTCACCTCCTGGAGCCCTGCTAGACTGTGCCCGACATGTGGCCGACTTTTCTCGAGGG 379
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 Db 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTCTGAGGACCGCAAG 439
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 Db 440 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGGGGTGCAAAATAGGCATCATGACC 499
 QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 Db 500 TTCAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACTTTCAAA 559
 QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 Db 560 GCCTGGGAAGGGTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
 QY 181 LeuProLeuTyrGluValValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 Db 620 TTGCCCTGTATGAGGTGATGACTTACGAGAGGCATTTTCGTAATTTGGGACTTT 673

RESULT 2

AAC55307

ID AAC55307 standard; cDNA; 2440 BP.

XX AAC55307;

XX 05-FEB-2001 (first entry)

DE Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.
 XX Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; anti-allergic;
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.
 OS Mus musculus.

XX Key Location/Qualifiers
 XX CDS 93..689

FT /*tag= a

FT /product= "activation-induced cytidine deaminase"

XX WO200058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

XX 24-JUN-1999; 99JP-0178999.

XX 27-DEC-1999; 99JP-0371382.

XX (NIBS) JAPAN TOBACCO INC.

XX (HONJ/) HONJO T.

XX Honjo T, Muramatsu M;

XX

DR WPI; 2000-611715/58.
 DR P-PSDB; AAB24197.
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 PS Claim 3; Page 126-130; 174pp; Japanese.
 XX
 CC The present sequence encodes mouse activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX
 SQ Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 1 other;

Alignment Scores:
 Pred. No.: 7,86e-113 Length: 2440
 Score: 1008.00 Matches: 183
 Percent Similarity: 95.94% Conservative: 6
 Best Local Similarity: 92.89% Mismatches: 8
 Query Match: 92.83% Indels: 0
 DB: 21 Gaps: 0

US-09-966-880A-8 (1-198) x AAC55307 (1-2440)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTTP 20
 DB 93 ATGGACAGCCTTCTGATGAGCAAAAGAGTTTCTTACCATTTCAAAATGTCGCTGG 152
 QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
 DB 153 GCCAAGGAGCGCATGAGACCTACCTCTGCTACGTGCTGAAGAGGAGAGATAGTGCCACC 212
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 DB 213 TCCTGCTCACTGAGTTCGGCCACCTTCGCAACAAGTCTGGCTGCCACGTGGAATGTTG 272
 QY 61 PheLeuArgTyrIleSerAspTTPAspLeuAspProGlyArgCysTyrArgValThrTP 80
 DB 273 TTCCTACGCTACATCTCAGACTGGACCTGGACCGCGCGGTGTACCGCTCACCTGG 332
 QY 81 PheThrSerTTPSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 DB 333 TTCACCTCCTGGAGCCGCTGCTATGACTGTGCGCGCAGCTGGCTGAGTTCTGAGATGG 392
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 DB 393 AACCTAACCTCAGCTGAGGATTTTCACCGCGCGCTTACTCTGTGAAGACCGCAAG 452
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 DB 453 GCTGAGCTGAGGGCTCGGAGAGACTGCACCGCTGGGTCCAGATCGGGATCATGACC 512
 QY 141 PheLysAspTyrPheTyrCysTTPAsnThrPheValGluAsnHisGluArgThrPheLys 160
 DB 513 TTCAAGAGACTATTTTACTGCTGGAATACATTTGTAGAAAAATCGTGAAGAACTTTCAA 572
 QY 161 AlaTTPGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 DB 573 GCTGGGAGGGGCTACATGAAATTTCTGTCGGGCTAACAGACAACCTTCGGGGCATCCTT 632

QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGly 197
 DB 633 TTGCCCTTGTACGAAGTCGATGACTTGGGAGATGTCATTCGTATGTTGGGA 683
 RESULT 3
 AAC55314
 ID AAC55314 standard; DNA; 6564 BP.
 XX
 AC AAC55314;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.
 XX
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; antiallergic;
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200058480-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 28-MAR-2000; 2000WO-JP01918.
 XX
 PR 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX
 (NISR) JAPAN TOBACCO INC.
 (HONJ/) HONJO T.
 HonJo T, Muramatsu M;
 WPI; 2000-611715/58.
 XX
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 Claim 17; Page 145-150; 174pp; Japanese.
 PS
 XX The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents a genomic DNA sequence of human AID.
 XX
 SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 other;

Alignment Scores:
 Pred. No.: 1.01e-67 Length: 6564

Score: 644.50 Matches: 177
 Percent Similarity: 24.02% Conservative: 0
 Best Local Similarity: 24.02% Mismatches: 2
 Query Match: 59.35% Indels: 560
 DB: 21 Gaps: 2

US-09-966-880A-8 (1-198) x AAC55314 (1-6564)

QY	3	SerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTIPAlaLys	22	
Db	1063	AGCCTCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGGCTAAG	1122	
QY	23	GlyArgAlaGluThrTyrLeuCysTyrValValLysArgAspSerAlaThrSerPhe	42	
Db	1123	GGTCGGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTT	1182	
QY	43	SerLeuAspPheGlyTyrLeuArgAsnLys	52	
Db	1183	TCACCTGGACTTTGGTTATCTTCGCAATAA -GGTATCAATTAAAGTCAGCTTTGCAAGCAG	1241	
QY	52	-----	52	
Db	1242	TTTAATGGTCAACTGTGAGTGCCTTTTAGAGCCACCTGCTGATGATTACTTCCATCCTT	1301	
QY	52	-----	52	
Db	1302	TTTTGGCATTTGTGCTCTATCACATTCCTCAAAATCCTTTTTTTTATTTCTTTTCCATG	1361	
QY	52	-----	52	
Db	1362	TCCATGCACCCATATTAGACATGGCCCAAAATATGTGATTAATTCCTCCCCAGTAATGC	1421	
QY	52	-----	52	
Db	1422	TGGGCACCCTAATACCACTCCTCTTCAGTGCCAGAACAACTGCTCCAAAAGTGTTA	1481	
QY	52	-----	52	
Db	1482	CCAGCTTTCCCTCAGACTCGAATTGCCTTTGAGATTAATTAAGCTAAAGCAATTTTATA	1541	
QY	52	-----	52	
Db	1542	TGGGAGAATATTACAGCTTGTCCAAAGCAAAATTTTAAATGTGAAAAACAAATTTGTGC	1601	
QY	52	-----	52	
Db	1602	TTAAGCATTTTGAATTAAGGAAGAAGAAATTTGGGAAAAAATTAACGGTGGTTCAATT	1661	
QY	52	-----	52	
Db	1662	CTGTTTTCCAAATGATTTCTTTCCCTCCTACTCACATGGTGGTAGGCCAGTGAATACA	1721	
QY	52	-----	52	
Db	1722	TTCAACATGGTATCCCCAGAAAACCTCAGAGAAGCCTCGGCTGATGATTAATTAATGA	1781	
QY	52	-----	52	
Db	1782	TCTTTGGGCTACCCGAGAGAATTACATTTCCAAGAGACTTCTTCACCAAAATCCAGATGG	1841	
QY	52	-----	52	
Db	1842	GTTTACATAAACTTTGGCCCATGGGTATCTCCTCTCTCTAACACGCTGTGACGTGGG	1901	
QY	52	-----	52	
Db	1902	CTTGGTGAATCTCAGGAGACATCCGTTGGGTGAAGGTCAATCCTGTGGCTCGTTGTT	1961	
QY	52	-----	52	
Db	1962	GATGTTATATACCATGCAATTTCTTTTGCTACATTTGTTGTAATACATCCCAATCT	2021	
QY	52	-----	52	

Db	2022	CCTTCCTATTCCGGTGACATGACACATTTCTATTTCAGAAAGCCTTTGATTTTATCAAGCACT	2081	
QY	52	-----	52	
Db	2082	TTCAATTTACTTCTCATGGCAGTGCTATTACTTCTTACAATACCCATCTGTCTGCTTT	2141	
QY	52	-----	52	
Db	2142	ACCAAAATCTATTTCCCTTTTTCAGATCTCCCAAATGGTCTCATAACTGTCTGCTGCT	2201	
QY	52	-----	52	
Db	2202	CCACCTAGTGGTCCAGGTATATTTCCACAATGTTACATCAACAGGCACCTTCTAGCCATTT	2261	
QY	52	-----	52	
Db	2262	TCCTTCTCAAAAGGTGCAAAAGCAACTTTCATAAAACAAAAATTAATCTTCGGTGAGGTA	2321	
QY	52	-----	52	
Db	2322	GTGTGATGCTGCTTCTCCCAACTCAGGCACCTTCTCTCTCAATTCACAAAAACCCA	2381	
QY	52	-----	52	
Db	2382	TAGCCTTCCCTTCACTCTGCAGGACTAGTGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGT	2441	
QY	52	-----	52	
Db	2442	GCTCTTTTGAGCAAGTTGCTTAGCCCTCTCTGTAAACACAGGACAATAGTGCAGCATCC	2501	
QY	52	-----	52	
Db	2502	CCAAAGATCATTCGACGAGACAAATGACTAAGGCTACACAGCGCGCAATAAAAGTCAGTGA	2561	
QY	53	-----	53	
Db	2562	ATTTTAGCGTGGTCTCTCTCTCTCCAGAACGGCTGCCACGTGAATTTGCTCTTCCCTC	2621	
QY	63	ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThr	82	
Db	2622	CGCTACATCTCGGACTGGGACCTTAGACCTGGCCGCTGTACCGCTCCTACCTGGTTCCAC	2681	
QY	83	SerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnPro	102	
Db	2682	TCCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGGAACCC	2741	
QY	103	AsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGlu	122	
Db	2742	AACCTCAGTCTGAGGATCTTCCCGCGCCCTTACTTCTGTGAGGACCGCAAGGCTGAG	2801	
QY	123	ProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLys	142	
Db	2802	CCCGAGGGGCTCGCGGGCTGCACCGCGCGGGTGCATAATAGCCATCATGACCTTCAA	2860	
QY	142	-----	142	
Db	2861	AGGTCCGAAAGGGCCTTCCGCGAGCGCAGTGCAGACGCCGCATTCGGGATTCGGATG	2920	
QY	142	-----	142	
Db	2921	CGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGCGGGGATTTCTGGTTCA	2980	
QY	142	-----	142	
Db	2981	CCTCTGGAGCCGAAATTAAGATTAGAACAGAGAAAAAGAGTGAATGGCTCAGAGACAAG	3040	
QY	142	-----	142	
Db	3041	CCCCCGAGAAATGAGAAAAATGGGGCCAGGGTTCCTTCTTCCCTCGATTTGGAACCTG	3100	
QY	143	-----	143	

-----AspTyr 144
 |||||

Db 3101 AACTGCTCTTCAACCCCATATACCCGCGCTTTTTTCCCTTTTTTTTTTTTGAAGATTAT 3160
 QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
 Db 3161 TTTTACTGCTGGAATACITTTGTAGAAACACGAAAGAACITTTCAAGCCCTGGGAGGG 3220
 QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181
 Db 3221 CTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCTTTTG 3271

RESULT 4

AAC55339
 ID AAC55339 standard; DNA; 11204 BP.
 AC AAC55339;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.
 XX
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; antiallergic;
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

XX Homo sapiens.

OS WO200058480-A1.

PN 05-OCT-2000.

PD 28-MAR-2000; 2000WO-JP01918.

PF 29-MAR-1999; 93JP-0087192.

PR 24-JUN-1999; 93JP-0178999.

PR 27-DEC-1999; 93JP-0371382.

XX (NIBS) JAPAN TOBACCO INC.

PA (HONJ/) HONJO T.

PI Honjo T, Muramatsu M;

XX WPI; 2000-611715/58.

DR Nucleic acid encoding activation induced cytidine deaminase, useful as

XX a target for drug development for immune-related diseases including

PT allergies -

XX Claim 17; Page 163-170; 174pp; Japanese.

XX The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC dermatological, antiasthmatic, ophthalmological, anti-HIV and
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders. Immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents a genomic DNA sequence of human AID.

XX SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;
 Alignment Scores:
 Pred. No.: 2.18e-67 Length: 11204
 Score: 644.50 Matches: 177
 Percent Similarity: 24.02% Conservative: 0
 Best Local Similarity: 24.02% Mismatches: 2
 Query Match: 59.35% Indels: 560
 DB: 21 Gaps: 2
 US-09-966-880A-8 (1-198) x AAC55339 (1-11204)
 QY 3 SerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLys 22
 Db 6278 AGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCCTGGCTAAG 6337
 QY 23 GlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPhe 42
 Db 6338 GGTGGCGGTGAGACCTACCTGTCTAGTAGTGAAGAGCGGTGACAGTGCTACATCCTTT 6397
 QY 43 SerLeuAspPheGlyTyrLeuArgAsnLys----- 52
 Db 6398 TCACGTGACTTTGGTTATCTTCGCAATAA--GGTATCAATTAAAGTCAGCTTTCGAACGAG 6456
 QY 52 ----- 52
 Db 6457 TTTAATGGTCAACTGTGAGTGCTTTTAGAGCCACCTCCTGATGTTACTTCCATCCTT 6516
 QY 52 ----- 52
 Db 6517 TTTTGGCATTTGTGCTCTATCATCATCTCTCAAAATCCTTTTATTTTCTTTTCCATG 6576
 QY 52 ----- 52
 Db 6577 TCCATGCACCCATATTAGACATGGCCCAAAATATGTGATTATTCCTCCCACTAATGC 6636
 QY 52 ----- 52
 Db 6637 TGGGCACCCCTAATACCACTCTTCTTCTAGTGCCAAAGAACAACTGCTCCCAAACTGTTA 6696
 QY 52 ----- 52
 Db 6697 CCAGCTTTCCTCAGCATCTGAATTGCTTTGAGATTAAATTAAGCTAAAGCAATTTTATA 6756
 QY 52 ----- 52
 Db 6757 TGGGAGAATATTATCAGCTTGTCCAAGCAAAATTTTAAATGTGAAAAACAATAATGTGTC 6816
 QY 52 ----- 52
 Db 6817 TTAAGCATTTTTCAAAATTAAGGAAGAGAAATTTGGGAAAAAATTAACGGTGGTTCAATT 6876
 QY 52 ----- 52
 Db 6877 CTGTTTTCAAAATGATTTCTTTTCCCTCTACTCATATGGTGTGCTAGGCCAGTGAATACA 6936
 QY 52 ----- 52
 Db 6937 TTCAACATGGTGATCCCAAGAAAACTCAGAGAAAGCCTCGGCTGATGATTAAATAATTGA 6996
 QY 52 ----- 52
 Db 6997 TCTTTCGGCTACCCGAGAGAAATTAATTTTCAAGAGACTTCTTTCACCAAAATCCAGATGG 7056
 QY 52 ----- 52
 Db 7057 GTTTACATAAACTTCTGCCCATGGGTATCTCTCTTCTTCAACACGCTGTGACGCTCTGGG 7116
 QY 52 ----- 52
 Db 7117 CTTGGTGAATCTCAGGGAAGCATCCGTGGGTGGAAGGTATCGTCTGGCTCTGTTGTT 7176

QY 52 ----- 52
Db 7177 GATGGTTATATTACCATGCAATTTCTTTGCTTACATTTTGATTGAATACATCCCAATCT 7236
QY 52 ----- 52
Db 7237 CCTTCTCTATTTCGGTGACATGACACATTTCTATTTCAGAAGGCTTTGATTTTATCAAGCACT 7296
QY 52 ----- 52
Db 7297 TTCAATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCACATCTGTCTGCTTT 7356
QY 52 ----- 52
Db 7357 ACCAAATCTATTTCCTCTTTTCAGATCTCTCCCAATGTCCTCATATAAATGTCCTGCCT 7416
QY 52 ----- 52
Db 7417 CCACCTAGTGTCCAGGTATATTTCACAAATGTTACATCAACAGGCATCTTAGCCATTT 7476
QY 52 ----- 52
Db 7477 TCCTTCTCAAAAGGTGCAAAAGCAACTTTCATAAACACAAATTAATCTTCGGTGAGGTA 7536
QY 52 ----- 52
Db 7537 GTGTGATGCTTCTCTCCAACTCAGCGCACTTCTCTTCTCTCAATTCACAAAAACCCA 7596
QY 52 ----- 52
Db 7597 TAGCCTCTCTTCACTCTGCAGGACTAGTCTGCCAAGGTTACGCTCTACCTACTGGTGT 7656
QY 52 ----- 52
Db 7657 GCTCTTTTGACCAAGTTGCTTAGCCCTCTCTGTAAACACAGGACAATAGTGAAGCATCC 7716
QY 52 ----- 52
Db 7717 CCAAAGATCATTCGAGGAGACAATGACTAAGGCTACCAGCGCGCAATAAAAGTCAGTGA 7776
QY 53 ----- AsnGlyCysHisValGluLeuPheLeu 62
Db 7777 ATTTTAGCTGGTCTCTCTCTCTCTCCAGAACGGCTGCCAGTGGAAATGCTCTTCCTC 7836
QY 63 ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThr 82
Db 7837 CGCTACATCTCGACTGGACCTAGACCTGCCGCTGTACCGCTCACCTGGTTCACC 7896
QY 83 SerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnPro 102
Db 7897 TCCTGGAGCCCTGCTTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGGGAACCC 7956
QY 103 AsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGlu 122
Db 7957 AACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAG 8016
QY 123 ProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLys 142
Db 8017 CCGAGGGCTGGGGGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTCAA- 8075
QY 142 ----- 142
Db 8076 AGGTGCGAAAGGGCTTTCGCGCAGCGGCGAGTGCAGACGCCGCAATTCGGGATTGGGATG 8135
QY 142 ----- 142
Db 8136 CGGAATGAATGAGTTAGTGGGAAGCTCGAGGGGAAGAAGTGGGGGGGATTCTGTTCA 8195
QY 142 ----- 142
Db 8196 CCTCTGGACCCGAAATTAAGATTAGAACGAGAAAAAGAGTGAATGGCTCAGACACAG 8255
QY 142 ----- 142

Db 8256 GCCCCGAGGAAATGAGAAAATGGGCCAGGGTTGCTTCTTCCCTCGATTGGAACCTG 8315
QY 143 ----- AspTyr 144
Db 8316 AACTGTCTTCTACCCCATATCCCCGCTTTTTCCTCTTTTTCCTTTTTCGAAGATTAT 8375
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db 8376 TTTTACTGCTGGAATACTTTTGTAGAAAACCAAGAAAGAACTTTCAAGCCCTGGGAAGG 8435
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181
Db 8436 CTGCATGAAATTCAGTTCGTCTCTCCACAGACTTCGCGGCATCCTTTTG 8486
RESULT 5
AAC55317
ID AAC55317 standard; DNA: 271 BP.
AC AAC55317;
XX
DT 05-FEB-2001 (first entry)
DE Human activation-induced cytidine deaminase exon 3 SEQ ID NO:13.
XX
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW immune related disease; allergy; allergic disease; anti-allergic;
KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; rosen disease; Digorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
XX
OS Homo sapiens.
XX WO2000058480-A1.
XX
XX 05-OCT-2000.
XX
PF 28-MAR-2000; 2000WO-JP01918.
XX
PR 29-MAR-1999; 99JP-0087192.
PR 24-JUN-1999; 99JP-0178999.
PR 27-DEC-1999; 99JP-0371382.
XX
PA (NISR) JAPAN TOBACCO INC.
PA (HONJ/) HONJO T.
PI Honjo T, Muramatsu M;
XX
DR WPI; 2000-611715/58.
XX
PT Nucleic acid encoding activation induced cytidine deaminase, useful as
PT a target for drug development for immune-related diseases including
PT allergies -
XX
PS Claim 18; Page 151; 174pp; Japanese.
XX
CC The present invention describes an activation-induced cytidine deaminase
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and
CC dermatological activities, and can be used in gene therapy. AID
CC polynucleotides are useful in methods for identifying drugs for the
CC treatment of B cell associated immune system disorders, immunodeficiency
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, rosen
CC disease, Digorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class)

DT 14-FEB-2002 (first entry)
 XX Human polynucleotide SEQ ID NO 173.
 XX
 KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;
 KW immune disorder; bacterial infection; fungal infection; cancer; tumour;
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;
 KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerability;
 KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulatory;
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.
 OS Homo sapiens.
 XX
 XX US2001039335-A1.
 XX
 XX 08-NOV-2001.
 XX
 XX 04-DEC-2000; 2000US-0729674.
 XX
 PR 10-APR-1997; 97US-126425P.
 PR 04-DEC-1997; 97US-067454P.
 PR 20-DEC-1997; 97US-068379P.
 PR 02-JAN-1998; 98US-070346P.
 PR 07-JAN-1998; 98US-070643P.
 PR 08-JAN-1998; 98US-070755P.
 PR 13-JAN-1998; 98US-071304P.
 PR 22-JAN-1998; 98US-072134P.
 PR 30-JAN-1998; 98US-073095P.
 PR 18-FEB-1998; 98US-075038P.
 PR 30-MAR-2000; 2000US-0539330.
 PR 23-NOV-1998; 98US-0197886.
 XX
 XX (JACO/) JACOBS K.
 PA (MCCO/) MCCOY J M.
 PA (LAVA/) LAVALLIE E R.
 PA (COLL/) COLLINS-RACIE L A.
 PA (EVAN/) EVANS C.
 PA (MERB/) MERBERG D.
 PA (TREA/) TREACY M.
 PA (AGOS/) AGOSTINO M J.
 PA (STEI/) STEININGER R J.
 PA (SPAU/) SPAULDING V.
 PA (WONG/) WONG G G.
 PA (CLAR/) CLARK H.
 PA (FECH/) FECHTEL K.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark H, Fechtel K;
 XX
 XX WPI: 2002-040725/05.
 DR P-PSDB: ABB55784.
 XX
 XX New secreted proteins and encoding polynucleotides, useful in gene
 PT therapies, particularly for preventing or treating autoimmune
 PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,
 PT stroke or inflammations
 XX
 PS Disclosure: Page 310-311; 349pp; English.
 XX
 XX The invention relates to isolated polynucleotides (ABA90876-ABA90968 and
 CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and
 CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in
 CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
 CC are deposited with the American Type Culture Collection (ATCC) with
 CC accession number 98599. The polynucleotides and encoded polypeptides have
 CC cytostatic, anti-inflammatory, immunomodulatory, vulnerability,
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic
 CC and anti-inflammatory activity and acting as cytokine modulators,
 CC

CC haematopoiesis regulators, tissue growth modulators and/or cadherin
 CC suppressors. The polypeptides and polynucleotides are useful in gene
 CC therapies, particularly for preventing, treating or ameliorating any of
 CC the following diseases: immune deficiency and disorders; e.g. bacterial
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or
 CC osteoarthritis; central and peripheral nervous system diseases and
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or
 CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
 CC folliculosis.
 XX
 XX Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1,77e-37 Length: 1534
 Score: 390.00 Matches: 83
 Percent Similarity: 61.62% Conservatives: 31
 Best Local Similarity: 44.86% Mismatches: 59
 Query Match: 35.91% Indels: 12
 DB: 24 Gaps: 3
 US-09-966-880A-8 (1-198) x ABA90962 (1-1534)
 QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25
 Db 639 ATGGATCCACCCACATTCCTTTCAACATTAACATGACCTTTGGGTCAGAGGACGGCAT 698
 QY 26 GluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45
 Db 699 GAGACTTACCTGTGTATGAGGTGGAGCGCATGCACATGACACCTCGGTGCTGCTGAAC 758
 QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55
 Db 759 CAGCGCAGGGCTTTCTCTGCAACACGCTCCACATAACACGCTTTCTTGAAGGCCGC 818
 QY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75
 Db 819 CATGCAGAGCTGTCTCTCTGGACGTGATTCCCTTTTGGAGGCTGGACCTGGACCGAC 878
 QY 76 TyrArgValThrTrpPheThrSerTrpSerProCystTyrAspCysAlaArgHisValAla 95
 Db 879 TACAGGGTTACCTGTCTGCTTACCTCTCGAGGCCCTGCTTACGTGTGCCCGAGGAATGGCT 938
 QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115
 Db 939 AAATTCATTTCAAAAACAAACACAGCTGAGCCTGTGCATCTTCACTGCCCGCATCTAT--- 995
 QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135
 Db 996 ---GATGATCAAGGAAGATGTGAGGAGGGCTCGCACCCCTCGCCGAGGCTGGGCCAAA 1052
 QY 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155
 Db 1053 ATTTCATTAATGACATACAGTGAATTAAGCATCTGCTGGGACCTTTGGGGACCCAG 1112
 QY 156 GluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
 Db 1113 GGATGTCCTTCCAGCCCTGGGATGATGATGACACACGCCCAAGACCTGAGTGGGAGG 1172
 QY 176 LeuArgArgIleLeu 180
 Db 1173 CTGCGGCCATCTC 1187
 RESULT 9
 AAA12409
 ID AAA12409 standard; cDNA; 1446 BP.
 XX
 AC AAA12409;

```
XX 25-JUL-2000 (first entry)
XX cDNA encoding a human RNA-associated protein.
XX
XX Human; RNA-associated protein; cell proliferation; cancer; inflammation;
XX immune response; reproductive disorder; actinic keratosis;
XX atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;
XX mixed connective tissue disease; myelofibrosis; primary thrombocythemia;
XX paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;
XX trauma; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 282..1446
XX /tag= a
XX /product= "RNA-associated protein"
XX
XX WO200015799-A2.
XX
XX 23-MAR-2000.
XX
XX 17-SEP-1999; 99NO-US21688.
XX
XX 17-SEP-1998; 98US-0156039.
XX 22-SEP-1998; 98US-0158720.
XX 04-NOV-1998; 98US-0186815.
XX 08-APR-1999; 99US-0128660.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;
XX Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J;
XX
XX WPI: 2000-271437/23.
XX P-PSDB; AAY84437.
XX
XX New polypeptides and polynucleotides, useful for preventing and
XX treating a disorder associated with increased or decreased expression
XX of RNA associated proteins -
XX
XX Claim 9; Page 119; 131pp; English.
XX
XX The present sequence encodes a human RNA-associated protein. The
XX expression of RNA-associated proteins is closely associated with
XX reproductive tissues, nervous tissues, cell proliferation including
XX cancer, inflammation and immune responses, and so they may be used
XX for diagnosis, treatment or prevention of cell proliferative,
XX immune/inflammatory disorders, and reproductive disorders. Diseases
XX and disorders which may be treated include actinic keratosis,
XX atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis,
XX mixed connective tissue disease, myelofibrosis, paroxysmal nocturnal
XX hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia
XX and cancers, and trauma.
XX
XX Sequence 1446 BP; 370 A; 374 C; 393 G; 308 T; 1 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,86e-37 Length: 1446
XX Score: 388.00 Matches: 82
XX Percent Similarity: 61.62% Conservative: 32
XX Best Local Similarity: 44.32% Mismatches: 59
XX Query Match: 35.73% Indels: 12
XX DB: 21 Gaps: 3
XX
XX US-09-966-880A-8 (1-198) x AAL12409 (1-1446)
XX
XX QY 6 MetAsnArgGlySerPheLeuTyrGlnPheLysAsnValArgTrrAlaLysGlyArgArg 25
XX |||:|:| ||| :|:| ||| ||| :|:|:|:|
XX Db 870 ATGGATCCACCAACATTCACCTTTCAACATTAACATGAACCTTGGTCAGAGCGGCAT 929
XX
XX QY 26 GluThrTyrLeuCysTyrValVallLysArgAspSerAlaThrSerPheSerLeuAsp 45
```

```
Db 930 GAGACTTACCTGTTATGAGGTGGAGCGATGCACAATGACACCTGGGTCTGCTGAAC 989
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 990 CAGCGCAGGGGCTTTCTATGCAACCCAGGCTCCACATAAACACGGTTTCCTTGAAGGCCGC 1049
QY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrrAspLeuAspProGlyArgCys 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 CATGCAGAGCTGTGCTTCTCTGGACGTGATTCCTTTTGGAACTGGACCTGGACCCAGGAC 1109
QY 76 TyrArgValThrTrrPheThrSerTrrSerProCysTyrAspCysAlaArgHisValAla 95
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 1110 TACAGGGTTACCTGCTTCACTTCCCTCGGAGCCCTGCTTCAGCTGTGCCCGAAATGGCT 1169
QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 1170 AAATTCATTTCAAAAAACAAACACGTCGACCTTTCACCTGCCCGCATCATAT--- 1226
QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 1227 ---GATGATCAAGGAAGATGTCAGGAGGGCTGCCGACCTGGCCGAGCTGGGCCAAA 1283
QY 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrrAsnThrPheValGluAsnHis 155
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 1284 ATTTCAATACTGACATACAGTGAATTTAAGCACTGCTGGACACCTTTGTGGACCAACAG 1343
QY 156 GluArgThrPheLysAlaTrrPgluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
|||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 1344 GGATGTCCCTCCAGCCCTGGGATGGACTAGAGGAGCAGACCCCAAGCCCTGAGTGGGAGG 1403
QY 176 LeuArgArgIleLeu 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1404 CTGCGGGGCATTCTG 1418
XX
XX RESULT 10
XX AAV48231
XX ID AAV48231 standard; cDNA; 610 BP.
XX
XX AC AAV48231;
XX
XX DT 09-NOV-1998 (first entry)
XX
XX DE Human RNA editing enzyme nucleotide sequence.
XX
XX KW ss; human; RNA editing enzyme; REE; pharmaceutical carrier; cancer;
XX viral disease; circulatory system disorder; RNA processing;
XX hypercholesterolaemia; alpha-galactosidase; apolipoprotein B.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT CDS 1..573
XX FT /*tag= a
XX FT /product= "RNA editing enzyme"
XX
XX US5804185-A.
XX
XX PD 08-SEP-1998.
XX
XX PF 13-MAR-1997; 97US-0816241.
XX
XX PR 13-MAR-1997; 97US-0816241.
XX
XX PA (INCY-) INCYTE PHARM INC.
XX
XX PI Bandman O, Goli SK;
XX
XX DR WPI: 1998-505585/43.
XX P-PSDB; AAW77092.
XX
XX PT Human RNA editing enzyme and polynucleotide(s) encoding it - useful
XX for recombinant production of the enzyme and treatment and detection
```

PT of disorders associated with incorrect RNA processing

XX Claim 4; Fig 1; 27pp; English.

XX The human RNA editing enzyme (REE) is used in a pharmaceutical carrier
XX for the treatment of cancer, viral diseases and circulatory system
XX disorders. The enzyme is used in vivo for the correct processing of RNA
XX transcripts of genes e.g. change of a codon in apolipoprotein B (apoB)
XX RNA to give a 100 and 48 kDa product transcribed from the same
XX gene. Certain disorders have been linked to incorrect RNA editing, e.g.
XX failure of apo B editing leads to excessive apoB 100 production and
XX hypercholesterolaemia. Other disorders thought to be linked to incorrect
XX RNA processing include aberrant alpha-galactosidase processing in
XX Fabry's disease and neurofibromatosis type I. The enzyme can be produced
XX recombinantly to treat related disorders. It can also be used to raise
XX antibodies for immuno-based detection of REE expression levels
XX e.g. ELISA.

XX Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 other;

Alignment Scores:

Pred. No.: 1-52e-35 Length: 610
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservativeness: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 19 Gaps: 4

US-09-966-880A-8 (1-198) x AAV48231 (1-610)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
DB 49 TTCTACTTCCAAATTTAAAAACCTATGGGAAGCCAGCATCGGAACAACTTGCTGCTGC 108
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
DB 109 TTCACCGTGAAGGATATAAAGCGCGCTCAGTTGTCTCCTGGAAGACG-----GGCGTC 162
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuPheLeuArgTyr 64
DB 163 TTCGGAACAGGTGATTCAGACCCATTGTCATGCAGAAAGGTGCTCTCTCTTGG 222
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
DB 223 TTCTCGAGGACATCTGTCTCTTAACACAAAGTACCAGGTACCTGTGTACACATCTTGG 282
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
DB 283 ACCCTTGGCCAGACTGTGCAGGGAGGTGGCGGAGTTCCTGCCAGGCACAGCAAGTG 342
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
DB 343 ATCTCACCATTCTCACCGCCGCTCTACTACTTCCAG--TATCCATGTTACCAAGGAG 399
QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
DB 400 GGGCTCCGAGCCTGAGTCAGGAAGGGGTGCTGTGGAGATCATGGACTATGAAGATTTT 459
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
DB 460 AAATATGTTGGGAACAACTTTCTGTACAAATGATATATGACCCATTCAAGCCTTGGGAAGGA 519
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 520 TTTAAACCAACTTTCGACTTCTGAAAGAGGCTACGGGAGAGTCTC 567

RESULT 11

AAV72058

ID AAAV72058 standard; cDNA; 610 BP.

AC AAAV72058;

XX 24-NOV-2000 (first entry)

XX cDNA encoding human RNA editing enzyme REE-2.

XX RNA editing enzyme; REE-2; human; HEPR homologue; REPR homologue;
XX phorbollin I homologue; cancer; tumour; autoimmune disorder;
XX circulatory system disorder; hypercholesterolaemia; viral infection;
XX neurological disease; neurofibromatosis; transcript editing; detection;
XX ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..573

XX FT /*tag= a

XX FT /product= "Human REE-2"

XX US6087108-A.

XX 11-JUL-2000.

XX 03-AUG-1998; 98US-0128395.

XX 13-MAR-1997; 97US-0816241.

XX (INCY-) INCYTE PHARM INC.

XX Goli SK, Bandman O;

XX WPI; 2000-531340/48.

XX P-PSDB; AAB11973.

XX Detecting polynucleotide encoding human RNA editing enzyme comprising
XX hybridizing an isolated and purified polynucleotide complementary to
XX the polynucleotide and detecting the hybridization complex -
XX Claim 5; Fig 1A-B; 27pp; English.

XX This sequence represents the cDNA encoding human RNA editing enzyme
XX REE-2. cDNA encoding REE-2 was initially isolated in a prostate tumour
XX cDNA library, with the present sequence representing a consensus. REE-2
XX has chemical and structural homology with the human apoB mRNA editing
XX protein HEPR (28% identity), the rat HEPR homologue REPR (30% identity),
XX and a portion of the mRNA editing enzyme phorbollin I (43% identity).
XX REE-2 was found to be expressed in a variety of cDNA libraries, a high
XX proportion of which were derived from tumours, neuronal tissues, immune
XX system cells or synovial tissue from arthritis patients. REE-2 is
XX therefore thought to be associated with the development of cancer,
XX autoimmune disorders, circulatory system disorders (e.g.
XX hypercholesterolaemia), viral infections and neurological diseases (e.g.,
XX neurofibromatosis). REE-2 or its nucleic acids may be used in the
XX diagnosis, treatment and prevention of such diseases via the modulation
XX of transcript editing, which in turn has effects on the encoded protein
XX (e.g., an alteration in protein activity). The invention specifically
XX relates to methods of detecting nucleic acids encoding human REE-2 in a
XX biological sample.

XX Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 other;

Alignment Scores:

Pred. No.: 1-52e-35 Length: 610
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservativeness: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 21 Gaps: 4

US-09-966-880A-8 (1-198) x AAA72058 (1-610)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30

DB 49 TTCTACTTCCAAATTTAAAAACCTATGGGAAGCCAGCATCGGAACAACTTGCTGCTGC 108

QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48

```

Db      109  TTACCGTGAAGGTATAAGCGCCGCTCAGTTGTCTCTGGAAGACG-----GCGGTC 162
QY      49  LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db      163  TTCCGAACACCGAGGTGATCTGAGACCCATTGTCATGCAGAAAGGTGCTTCCTCTCTGG 222
QY      65  IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
Db      223  TTCTGCGAGACATATCTCTCTTAACACAAAGTACCAAGTCCAGTGCATGATCATCTTGG 282
QY      85  SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db      283  AGCCCTTGCCAGACTGTCAGGGAGGTGGCGAGTTCCTGCCAGGCACACCAACGTG 342
QY      105  SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
Db      343  AATCTCACCATCTTCACCGCCGCTCTACTACTTCCAG---TATCATGTTACCAAGAG 399
QY      125  GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db      400  GGGCTCCGACGCTCAGTCAGGAAGGGTCGCTGTGGAGATCATGGACTATGAAGATTTT 459
QY      145  PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db      460  AAATATCTTGGAAACTTTGTGTACATGATATGAGCCATTCAAGCCTTGAAGGGA 519
QY      165  LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      520  TTAACCAACTTTCGACTTCTGAAAGAGGCTACGGGAGAGTCTC 567

RESULT 12
ID      AAS41420 standard; cdna; 950 BP.
XX
AC      AAS41420;
XX
DT      17-DEC-2001 (first entry)
XX
DE      cdna encoding novel human enzyme polypeptide #636.
KW      Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW      ligase; hyperproliferative disorder; immunodeficiency disorder;
KW      autoimmune disorder; neurological disorder; metabolic disorder;
KW      inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW      blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW      anti arthritic; nephrotropic; anticoagulant; ss.
OS      Homo sapiens.
XX
PN      W0200155301-A2.
XX
PD      02-AUG-2001.
XX
PF      17-JAN-2001; 2001WO-US01239.
XX
PR      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      24-FEB-2000; 2000US-0184664.
PR      02-MAR-2000; 2000US-0186350.
PR      16-MAR-2000; 2000US-0189874.
PR      17-MAR-2000; 2000US-0190076.
PR      18-APR-2000; 2000US-0198123.
PR      19-MAY-2000; 2000US-0205515.
PR      27-JUN-2000; 2000US-0209467.
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      14-JUL-2000; 2000US-0217496.
PR      01-NOV-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      26-JUL-2000; 2000US-0220964.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225266.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226681.
PR      22-AUG-2000; 2000US-0226868.
PR      22-AUG-2000; 2000US-0227182.
PR      23-AUG-2000; 2000US-0227009.
PR      30-AUG-2000; 2000US-0228924.
PR      01-SEP-2000; 2000US-0229287.
PR      01-SEP-2000; 2000US-0229343.
PR      01-SEP-2000; 2000US-0229344.
PR      01-SEP-2000; 2000US-0229345.
PR      05-SEP-2000; 2000US-0229509.
PR      05-SEP-2000; 2000US-0229513.
PR      06-SEP-2000; 2000US-0230437.
PR      06-SEP-2000; 2000US-0230438.
PR      08-SEP-2000; 2000US-0231242.
PR      08-SEP-2000; 2000US-0231243.
PR      08-SEP-2000; 2000US-0231244.
PR      08-SEP-2000; 2000US-0231413.
PR      08-SEP-2000; 2000US-0231414.
PR      08-SEP-2000; 2000US-0232080.
PR      08-SEP-2000; 2000US-0232081.
PR      12-SEP-2000; 2000US-0231968.
PR      14-SEP-2000; 2000US-0232397.
PR      14-SEP-2000; 2000US-0232398.
PR      14-SEP-2000; 2000US-0232399.
PR      14-SEP-2000; 2000US-0232400.
PR      14-SEP-2000; 2000US-0232401.
PR      14-SEP-2000; 2000US-0233063.
PR      14-SEP-2000; 2000US-0233064.
PR      21-SEP-2000; 2000US-0233065.
PR      21-SEP-2000; 2000US-0234223.
PR      21-SEP-2000; 2000US-0234274.
PR      25-SEP-2000; 2000US-0234997.
PR      25-SEP-2000; 2000US-0234998.
PR      26-SEP-2000; 2000US-0235484.
PR      27-SEP-2000; 2000US-0235834.
PR      27-SEP-2000; 2000US-0235836.
PR      29-SEP-2000; 2000US-0236327.
PR      29-SEP-2000; 2000US-0236367.
PR      29-SEP-2000; 2000US-0236368.
PR      29-SEP-2000; 2000US-0236369.
PR      29-SEP-2000; 2000US-0236370.
PR      02-OCT-2000; 2000US-0236802.
PR      02-OCT-2000; 2000US-0237037.
PR      02-OCT-2000; 2000US-0237038.
PR      02-OCT-2000; 2000US-0237039.
PR      02-OCT-2000; 2000US-0237040.
PR      13-OCT-2000; 2000US-0239935.
PR      13-OCT-2000; 2000US-0239937.
PR      20-OCT-2000; 2000US-0240960.
PR      20-OCT-2000; 2000US-0241221.
PR      20-OCT-2000; 2000US-0241785.
PR      20-OCT-2000; 2000US-0241786.
PR      20-OCT-2000; 2000US-0241787.
PR      20-OCT-2000; 2000US-0241808.
PR      20-OCT-2000; 2000US-0241809.
PR      01-NOV-2000; 2000US-0241826.
PR      01-NOV-2000; 2000US-0244617.
PR      08-NOV-2000; 2000US-0246474.
PR      08-NOV-2000; 2000US-0246475.

```

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
DR
DR WPI: 2001-465566/50.
DR P-PSDB; AU233550.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases .
XX
PS Claim 4; SEQ ID No 646; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders

CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 950 BP; 219 A; 289 C; 219 G; 220 T; 3 other;

Alignment Scores:
Pred. No.: 2.88e-35 Length: 950
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 22 Gaps: 4

US-09-966-880A-8 (1-198) x AAS41420 (1-950)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrrAlaLysGlyArgGluThrTyrLeuCys 30
Db 110 TTCTACTTCCAAATTTAAACCTATGGGAAGCCACAGATCGGAACAACTTGGCTGTGC 169
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
Db 170 TTCACCGTGGAGGTATAAAGCCGCTCAGTTGTCTCTCTGGAGAGCG-----GGCGTC 223
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db 224 TTCCGAAACGAGGTGGATCTTGAGACCCATTGTATGCAGAAAGGTCTTCTCTCTCTGG 283
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
Db 284 TTCTGCGAGCATATCTCTCTCTAACACAAAGTACCAGTCCCTGGTACACATCTGG 343
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db 344 AGCCCTTGCCAGACTGTGCAGGGGAGGTGGCCGAGTTCTTGGCCAGGCACAGCAGTG 403
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
Db 404 AATCTCACCATCTTCCACGCCGCTCTACTACTTCCAG---TATCCATGTTACAGGAG 460
QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db 461 GGGCTCCGAGCCTGAGTCAGGAAGGGGTGCTGTGGAGATCATGGACTATGAAGATTTT 520
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db 521 AAATATTGTTGGAAAACTTTGTGTACATGATGATGAGCCATTCAAGCCTTGAAGGGA 580
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 581 TTAATAACCAACTTTCGACTTCTGAAAGAGGCTACGGGAGAGTCCTC 628

RESULT 13
AAFI6264
ID AAF16264 standard; cDNA; 987 BP.
XX
XX AAF16264;
XX
XX 13-MAR-2001 (first entry)
XX
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:699.
DE
XX Human prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerrary; gastrointestinal; nephrotropic; antinefactive; gynaecological;
KW antibacterial; gene therapy; immune; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX

```
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
XX
PD
PD
XX
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR P-PSDB; AAB57061.
XX
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX
PS Claim 1; Page 1129; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 987 BP; 231 A; 298 C; 232 G; 224 T; 2 other;
XX
Alignment Scores:
Pred. No.: 3,048-35 Length: 987
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 21 Gaps: 4
US-09-966-880A-8 (1-198) x AAF16264 (1-987)
QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
147 TTCTACTCCATTAAACACCTATGGGAGCCACGATCGGAACCAACTGGCTGTGC 206
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
DB 207 TTCACCGTGAAGGTATAAAGCGCGCTGCTCTCTGGAAGACG-----GGCGTC 260
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuPheLeuArgTyr 64
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
DB 261 TTCCGAACACAGGTGATCTGAGACCCATTGTCATGCAGAAAGGTGCTCTCTCTTGG 320
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
DB 321 TTCTCGGACGACATCTGCTCCTAACACAAAGTACAGGTACCTGTGTACATCTTGG 380
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
QY 85 SerProCysTyrAspCysAlaAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
DB 381 AGCCCTTGCCAGACTGTGAGGGGAGGTGGCCGAGTTCCTGGCCAGGCACAGCAACGTG 440
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluArgArgLysAlaGluProGlu 124
||| :|||:|||||:|||||: ||| ||| |||:|||||:|||||:
```

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-465566/50.
DR P-PSDB; AAU23537.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
XX Claim 4; SEQ ID No 633; 1180pp; English.
PS
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 987 BP; 231 A; 298 C; 232 G; 224 T; 2 other;

Alignment Scores:

Pred. No.:	3 04e-35	Length:	987
Score:	369.50	Matches:	79
Percent Similarity:	58.52%	Conservative:	24
Best Local Similarity:	44.89%	Mismatches:	64
Query Match:	34.02%	Indels:	9
DB:	22	Gaps:	4

US-09-966-880A-8 (1-198) x AAS41407 (1-987)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArgGluThrTyrLeuCys 30
 DB 147 TTCTACTTCCAAATTTAAACACCTTATGGGAAGCCAGCATCGGAACAACTTGGCTGTGC 206

QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
 DB 207 TTCACCGTGAAGGTATAAAGCGCGCTCAGTTGTCTCTGGAAGACG-----GGCGTC 260

QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuPheLeuArgTyr 64
 DB 261 TTCGGAACAGGTGATCTCAGAGCCCATTTGTCATGCAGAAAGGTGCTTCTCTCTTGG 320

QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
 DB 321 TTCTGCGAGCATACTGTCTCTTAACACAAAGTACCAGGTCCCTGGTACACATCTTGG 380

QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
 DB 381 AGCCCTTCCAGAGCTGCGAGGAGGTGGCGGAGTTCTCTGGCCAGGCACGCAACGTG 440

QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
 DB 441 AATCTCAGCATCTTCCAGCGCGCTCTACTACTTCCAG---TATCCATGTTACCGAG 497

QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
 DB 498 GGGTCCCGACGCTGATGTCAGGAAGGGTGGCTGAGATCATGGACTATGAAGATT 557

QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
 DB 558 AAATATTGTTGGGAAACTTGTGTACATGATATGAGCCATCAAGCCTTGAAGGGA 617

QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 DB 618 TTAACCAACCACTTTCGACTTCTGAAAGAAAGGCTACGGGAGAGTCTC 665

RESULT 15
 AAI59847/c
 ID AAI59847 standard; cDNA; 1120 BP.
 AC AAI59847;
 XX
 XX
 XX 22-OCT-2001 (first entry)
 XX
 XX Human polynucleotide SEQ ID NO 3836.
 XX
 KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200153312-A1.
 XX
 XX 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX

(HYSE-) HYSEQ INC.

PA XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR P-PSDB; AAM40691.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 3836; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX

Sequence 1120 BP; 250 A; 268 C; 332 G; 270 T; 0 other;

Alignment Scores:

Pred. No.: 3,64e-35 Length: 1120
 Score: 369.50 Matches: 79
 Percent Similarity: 58.52% Conservative: 24
 Best Local Similarity: 44.89% Mismatches: 64
 Query Match: 34.02% Indels: 9
 DB: 22 Gaps: 4

US-09-966-880A-8 (1-198) x AAI59847 (1-1120)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArgGluThrTyrLeuCys 30
 DB 977 TTCTACTTCCAAATTTAAACACCTATGGGAGCCAGCATCGGAACAACTTGGCTGTGC 918

QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
 DB 917 TTCACCGTGAAGGTATAAAGCGCGCTCAGTTGTCTCTGGAAGACG-----GGCGTC 864

QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuPheLeuArgTyr 64
 DB 863 TTCGGAACAGGTGATCTCAGAGCCCATTTGTCATGCAGAAAGGTGCTTCTCTCTTGG 804

QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
 DB 803 TTCTGCGAGCATACTGTCTCTTAACACAAAGTACCAGGTCCCTGGTACACATCTTGG 744

QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
 DB 743 AGCCCTTCCAGAGCTGTCAGGAGGAGTGGCGGAGTTCTCTGGCCAGGCACGCAACGTG 684

QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
 DB 683 AATCTCAGCATCTTCCAGCGCGCTCTACTACTTCCAG---TATCCATGTTACCGAGG 627

QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
 DB 626 GGGTCCCGACGCTGATGTCAGGAAGGGTGGCTGAGATCATGGACTATGAAGATT 567

QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
 DB 558 AAATATTGTTGGGAAACTTGTGTACATGATATGAGCCATCAAGCCTTGAAGGGA 617

Db 566 AAATATTCTTGGGAAACTTTGTTGTACATGATATGAGCCATTCAAGCCTTGAAGGGA 507

Qy 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
||| ||| ||| |||:::||||| |||
||| ||| ||| |||:::||||| |||

Db 506 TTAATAACCAACTTTCGACTTCTGAAAGAGAGGCTACGGGAGAGTCTC 459

Search completed: June 19, 2003, 17:58:08
Job time : 233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 19, 2003, 13:52:31 ; Search time 1384 Seconds
(without alignments)
2316.984 Million cell updates/sec

Title: US-09-966-880A-8
Perfect score: 1086
Sequence: 1 MDSLMMNRKFLYQFKNVRW.....ILLPLYEVDLDAFTLGL 198

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09966880/runat_14062003_175645_16282/app_query.fasta_1.391
-DB=EST -Qfmt=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 @cgn_1.1906 @runat_14062003_175645_16282 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1086	100.0	743	12	BG686133	BG686133 602638412
2	1086	100.0	856	9	AL559877	AL559877 602638412
3	1086	100.0	872	12	BG758510	BG758510 6027112721
4	1086	100.0	953	14	BQ065440	BQ065440 AGENCOURT
5	1086	100.0	1052	14	BQ055935	BQ055935 AGENCOURT
6	1062	97.8	820	12	BG757089	BG757089 6027115124
7	1052	96.9	693	12	BG757392	BG757392 6027110222
8	1033	95.1	942	12	BF975166	BF975166 6022444657
9	986	90.8	757	9	AJ446140	AJ446140 AJ446140
10	977	90.0	696	9	AJ453647	AJ453647 AJ453647
11	910	83.8	729	9	AJ450296	AJ450296 AJ450296
12	884	81.4	688	9	AJ450317	AJ450317 AJ450317
13	868	79.9	522	12	BG144705	BG144705 ut73f07.y
14	858	79.0	541	12	BF238155	BF238155 601811880
15	850	78.3	650	9	AJ449745	AJ449745 AJ449745
16	780	71.8	889	12	BG686876	BG686876 602650861
17	724	66.7	623	9	AJ450295	AJ450295 AJ450295
18	675	62.2	693	12	BF975096	BF975096 6022445679
19	665	61.2	653	10	BG637360	BG637360 BB637360
20	378	34.8	884	13	B1834760	B1834760 603090364
21	371.5	34.2	986	14	BQ877348	BQ877348 AGENCOURT
22	370.5	34.1	1100	14	BM914942	BM914942 AGENCOURT
23	370	34.1	1073	12	BG342101	BG342101 602463048
24	369.5	34.0	741	12	BE886229	BE886229 601509806
25	369.5	34.0	923	12	BG025943	BG025943 602292096
26	369.5	34.0	940	12	BG032405	BG032405 602301392
27	369.5	34.0	974	14	BQ707723	BQ707723 AGENCOURT
28	369.5	34.0	1005	13	BM474553	BM474553 AGENCOURT
29	369.5	34.0	1020	13	BM424095	BM424095 AGENCOURT
30	369.5	34.0	1063	14	BM809617	BM809617 AGENCOURT
31	369.5	34.0	1117	14	BM917194	BM917194 AGENCOURT
32	369.5	34.0	1119	11	BC021080	BC021080 Homo sapi
33	369.5	34.0	1128	13	BM559604	BM559604 AGENCOURT
34	369	34.0	1549	11	BC009683	BC009683 Homo sapi
35	367.5	33.8	782	12	BE882870	BE882870 601509274
36	367.5	33.8	1665	13	BM473129	BM473129 AGENCOURT
37	363.5	33.5	1005	14	BQ056030	BQ056030 AGENCOURT
38	363.5	33.5	1052	14	BQ063928	BQ063928 AGENCOURT
39	363.5	33.5	1083	14	BQ058970	BQ058970 AGENCOURT
40	362.5	33.4	1021	14	BQ061901	BQ061901 AGENCOURT
41	360.5	33.2	772	14	BM935138	BM935138 UI-M-BH3 -
42	357	32.9	1020	14	BQ052514	BQ052514 AGENCOURT
43	356.5	32.8	787	10	BE270707	BE270707 601185701
44	356.5	32.8	1021	14	BQ065295	BQ065295 AGENCOURT
45	356	32.8	836	12	BG744603	BG744603 602722683

ALIGNMENTS

RESULT 1
BG686133
LOCUS 602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5',
DEFINITION mRNA sequence.
ACCESSION BG686133
VERSION BG686133.1 GI:13917530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 743)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1626 row: 9 column: 03
High quality sequence stop: 740.
FEATURES
source
1. 743
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 176 a 197 c 188 g 182 t
ORIGIN

Alignment Scores:

Pred. No.: 7,26e-131 Length: 743
Score: 1086.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BG686133 (1-743)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 78 ATGGACACCCCTCTGTATGTAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 137
Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 138 GCTAAGGGTCGGCTGAGACCTACCTGCTACGTAGTGAAGGCGGTGACAGTGTCTACA 197
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 198 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAAGAACCGCTGCCACGTGGAATTCCTC 257
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 258 TTCTCGCGTACATCTCGGACTGGACCTAGACCCCTGGCGCTGCTACCGCGTCACTGG 317
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 318 TTCACCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 377
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 378 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 437
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValcIleAlaIleMetThr 140
Db 438 GCTGAGCCGAGGGGCTGGCGGCGCTGCACCGCGCGGGTGCATAATGCCATCATGACC 497
Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 498 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACATTTCAA 557

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgAlaGluLeu 180
Db 558 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 617
Qy 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 618 TTCCCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGTAATTTGGACTT 671

RESULT 2

AL559877

LOCUS

DEFINITION

AL559877 LTI_FL011_BC1 Homo sapiens cDNA clone CS0DG003YB14 5 prime

mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 856

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DG003YB14"

/clone_lib="LTI_FL011_BC1"

/sex="male"

/tissue_type="B cells from Burkitt lymphoma"

/lab_host="DH10B"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-stranded cDNA was digested with Not I and cloned

into the Not I and Eco RV sites of the pCMVSPORT 6 vector.

Library was constructed by Life Technologies. Contact :

Feng Liang Life Technologies, a division of Invitrogen

9800 Medical Center Drive Rockville, Maryland 20850, USA

Fax : (1) 301 610 8371 Email : filiang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 209 a 217 c 202 g 226 t

ORIGIN

Alignment Scores:

Pred. No.: 9,22e-131 Length: 856

Score: 1086.00 Matches: 198

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AL559877 (1-856)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20

Db 19 ATGGACACCCCTCTGTATGTAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 78

Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

Db 79 GCTAAGGGTCGGCTGAGACCTACCTGCTACGTAGTGAAGGCGGTGACAGTGTCTACA 138

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

Db 139 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAAGAACCGCTGCCACGTGGAATTCCTC 198

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLCM2108 row: p column: 10

High quality sequence stop: 634.

FEATURES

Location/Qualifiers
source

1..953
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:592977"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 238 a 236 c 233 g 246 t

ORIGIN

Alignment Scores:
Pred. No.: 1..1e-130 Length: 953
Score: 1086.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-966-880A-8 (1-198) x BQ065440 (1-953)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
|||||
Db 64 ATGGACAGCCCTCTTGATGACCGGAGGAAGTTCTTACCAATTCACAAATGTCCGCTGG 123
Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
|||||
Db 124 GCTAAGGTCGGCGTGAGACCTACTGTGTACGTAGTAGTAGGAGCGGTGACAGTGCTACA 183
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|||||
Db 184 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGATTCGTC 243
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 244 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 303
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 304 TTCACCTCTGGAGCCCTCTGTAGCATGTGCCGACATGTGGCCGACTTCTCGGAGGG 363
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 364 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 424 GCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACC 483
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 484 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCAACCAAGAACTTTCAAA 543
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
|||||
Db 544 GCCTGGGAAGGCTGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 603
Qy 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
|||||

Db 604 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTCTGACTTTGGGACTT 657

RESULT 5

BQ055935

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ055935 1052 bp mRNA linear EST 29-MAR-2002
AGENCOURT_6796291 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5808181
5', mRNA sequence.

BQ055935

EST

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1052)

NIH-MGC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Lou Staudt

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM2051 row: m column: 14

High quality sequence stop: 665.

Location/Qualifiers

1..1052

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5808181"

/clone_lib="NIH_MGC_99"

/tissue_type="lymphoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGACGAG(G). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH_MGC

Library."

BASE COUNT 276 a 250 c 258 g 266 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 1.31e-130 Length: 1052

Score: 1086.00 Matches: 198

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 14 Gaps: 0

US-09-966-880A-8 (1-198) x BQ055935 (1-1052)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20

|||||

Db 64 ATGGACAGCCCTCTTGATGACCGGAGGAAGTTCTTACCAATTCACAAATGTCCGCTGG 123

Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

|||||

Db 124 GCTAAGGTCGGCGTGAGACCTTACCTGTGTACGTAGTAGGAGCGGTGACAGTGCTACA 183

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

|||||

Db 184 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGATTCGTC 243

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

|||||

```

Db      244  TTCTCCGCTACATCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTCGG 303
QY      81  PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAlaAspPheLeuArgGly 100
Db      304  TTCACCTCTCGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACACTTTCGCGAGGG 363
QY      101  AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      364  AACCCCAACCTCAGTCTGAGGATCTTCACGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
QY      121  AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      424  GCTGACCCGAGGGGCTGGCGGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 483
QY      141  PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      484  TTCAAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATTTCAAA 543
QY      161  AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      544  GCCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTT 603
QY      181  LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      604  TTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTGCTACTTTGGGACTT 657

RESULT 6
BG757089
LOCUS   6027115124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
DEFINITION
mRNA sequence.
ACCESSION
BG757089
VERSION
BG757089.1 GI:14067742
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 820)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1704 row: 0 column: 06
High quality sequence stop: 675.
FEATURES
Location/Qualifiers
1..820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4855517"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
210 a 230 c 200 g 180 t
BASE COUNT
ORIGIN

```

Alignment Scores:

```

Pred. No.: 1.17e-127 Length: 820
Score: 1062.00 Matches: 193
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.79% Indels: 0
DB: 12 Gaps: 0

```

US-09-966-880A-8 (1-198) x BG757089 (1-820)

```

QY      1  MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      85  ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATTTCAAAAATGTCCGCTGG 144
QY      21  AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      145  GCTTAGGGTCGGCTGAGACCTACCTGTGCTAGTAGTAGAAGAGCGCTGACATGCTGTACA 204
QY      41  SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      205  TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGCTGGAATTTGCTC 264
QY      61  PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      265  TTCTTCGCTTACATCTCGGACTCGGACTAGACCTCGCCGCTGCTACCGCTCACCTCGG 324
QY      81  PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      325  TTCACCTCTCGAGCCCTCTGCTAGGACTGTGCCGACATGTGGCCGACTTTTCTCGGAGGG 384
QY      101  AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      385  AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 444
QY      121  AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      445  GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 504
QY      141  PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      505  TTCAAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATTTTCAAA 564
QY      161  AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      565  GCCTGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCGCATCCTT 624
QY      181  LeuProLeuTyrGluValAspLeuArgAspAlaPhe 193
Db      625  TTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTC 663

```

RESULT 7

```

BG757392 602711022F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851580 5',
LOCUS
DEFINITION
mRNA sequence.
ACCESSION
BG757392
VERSION
BG757392.1 GI:14068045
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
1 (bases 1 to 693)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1694 row: k column: 05
 High quality sequence stop: 693.
 Location/Qualifiers

FEATURES

source
 1..693
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4851580"
 /clone_lib="NIH_MGC_48"
 /tissue_type="Primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 157 a 188 c 178 g 170 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.78e-126 Length: 693
 Score: 1052.00 Matches: 194
 Percent Similarity: 98.48% Conservative: 0
 Best Local Similarity: 98.48% Mismatches: 2
 Query Match: 96.87% Indels: 1
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BG757392 (1-693)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
 |||
 Db 81 ATGGACAGCCTCTGTATGATACCGGAGGAAATTTCTTACCAATTCAAATATGTCGGCTGG 140
 |||
 QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLys-ArgArgAspSerAlaTh 40
 |||
 Db 141 GCTAAGGGTGGGGTGAGACCTACCTGTGTAGTAGTAACGAGGCGTGACAGTGCTAC 200
 |||
 QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60
 |||
 Db 201 ATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACATGGAAATGCT 260
 |||
 QY 60 uPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTr 80
 |||
 Db 261 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTCAGCTG 320
 |||
 QY 80 pPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL 100
 |||
 Db 321 GTTACCTCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACTTCTTCTGCGAGG 380
 |||
 QY 100 YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
 |||
 Db 381 GAACCCCAACCTCAGTCTGAGGATCTCACCGCGCGCTCTACTTCTGTGAGGACCGCAA 440
 |||
 QY 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
 |||
 Db 441 GGCTGAGCCGAGGGGCTGGCGGCTGTGACCCGCGGGGTGCAATATGCCATCATGAC 500
 |||
 QY 140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
 |||
 Db 501 CTTCAAGATATTCCTTACTGCTGGAAATCTTTTGTAGAAACCATGAAAGAACTTTCAA 560
 |||
 QY 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
 |||
 Db 561 AGCCTGGGAGGGCTCATGAAATTCAGTTCGCTCTCCAGACACCTTCGGCGCATCCT 620
 |||
 QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrTr 196
 |||
 Db 621 TTTGCCCTGTTATGAGTTTGATGACTTAGGAGCGCATTTCTGCTACTTTG 669

RESULT 8

BF975166
 LOCUS 602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4335639 5',
 DEFINITION mRNA sequence.
 ACCESSION BF975166 942 bp mRNA linear EST 22-JAN-2001
 VERSION BF975166
 KEYWORDS BF975166.1 GI:12342381
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 942)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1207 row: a column: 16
 High quality sequence stop: 707.
 Location/Qualifiers

FEATURES

source
 1..942
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4335639"
 /clone_lib="NIH_MGC_48"
 /tissue_type="Primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 249 a 209 c 251 g 233 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9.04e-124 Length: 942
 Score: 1033.00 Matches: 194
 Percent Similarity: 97.00% Conservative: 0
 Best Local Similarity: 97.00% Mismatches: 4
 Query Match: 95.12% Indels: 2
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BF975166 (1-942)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGln-PhelYsAsnValArgTr 20
 |||
 Db 78 ATGGACAGCCTCTTGTATGATACCGGAGGAAATTTCTTACCAATTCAAATATGTCGGCTG 137
 |||
 QY 20 pAlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaTh 40
 |||
 Db 138 GGCTAAGGGTGGGGTGAGACCTACCTGTGTAGTAGTAAGAGGCGTGACAGTGCTAC 197
 |||
 QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60
 |||
 Db 198 ATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACATGGAAATGCT 257
 |||
 QY 60 uPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTr 80
 |||
 Db 258 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTCAGCTG 317


```
QY      80 pPhethrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL 100
      |||
Db      318 GTTACCTCTCGAGGCCCTGCTACGACTGTGCCGACATGTGGCGAGCTTCTCGGAGG 377
      |||
QY      100 yAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
      |||
Db      378 GAACCCCAACCTCAGTCTGAGATCTTCACCGCGCCCTCTACTTCTGTGAGGACGCAA 437
      |||
QY      120 salaGluProGluGlyLeuArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
      |||
Db      438 GCCTGAGCCCGAGGGCTGCGCGCTGCACGAGCCGGGTGCAAAATGACCATCATGAC 497
      |||
QY      140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
      |||
Db      498 CTTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAAGAACTTTCAA 557
      |||
QY      160 salaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
      |||
Db      558 AGCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCT 617
      |||
QY      180 uLeuProLeuTyrGlu-ValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
      |||
Db      618 TTGGGCCCTGTATGAGGGTTGATGACTTACGAGACCAATTCGTACTTGGGACTTG 673
      |||

RESULT 9
AJ446140
LOCUS      757 bp mRNA linear EST 19-APR-2002
DEFINITION      AJ446140 riken1 Gallus gallus cdna clone 12j9rl, mRNA sequence.
ACCESSION      AJ446140
VERSION      AJ446140.1 GI:20213361
KEYWORDS      EST.
SOURCE      chicken.
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES             source
1..757
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="12j9rl"
/clone_lib="riken1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT      180 a 205 c 184 g 188 t
ORIGIN
Alignment Scores:
Pred. No.:      8.58e-118 Length:      757
Score:      986.00 Matches:      178
Percent Similarity: 94.44% Conservative: 9
Best Local Similarity: 89.90% Mismatches: 11
Query Match:      90.73% Indels:      0
DB:      9 Gaps:      0

US-09-966-880A-8 (1-198) x AJ446140 (1-757)

QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
      |||
Db      97 ATGGACAGCCTCTTGATGAGAGGAGGAGCTCTTCCTCACTCAAAATTCAGAACTTCGCGCTG 156
      |||
QY      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
      |||
```

```
Db      157 GCCAAAGGCGTGTGTAACCTACCTCTGTATGTTGTGAAGCGCCGTGACAGTCTACA 216
      |||
QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
      |||
Db      217 TCATGCTCCCTGGACCTTTGGATACCTGCGTAAACAAGATGGTGTGCCATGTGGAGGTCTC 276
      |||
QY      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
      |||
Db      277 TTCTACGCTACATCTCAGCTTGGACCTGGACCCAGCCGCTGCTACCGCATCATCATGG 336
      |||
QY      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
      |||
Db      337 TTCACCTCTCGAGGCCCTGTTATGACTGTGCGCGACATGTGGCTGACTTCTCTCGTGCC 396
      |||
QY      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
      |||
Db      397 TACCAAACTTGACCTCCGCACTTTCACCTGCCCGCTCTACTTCTGTGAAGATCGCAAG 456
      |||
QY      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
      |||
Db      457 GCTGAGCCTCAGGGGCTGAGACGCTGCACCGGCTGGGGCCCAAAATCGCATCATGACT 516
      |||
QY      141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
      |||
Db      517 TTCAAAGATTCTTCTACTGCTGGAACACGTTTGTGGAGAACAAGGAAAAAGACATTCAAA 576
      |||
QY      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
      |||
Db      577 GCCTGGGAAGGGCTGCATGAAATCTCTGTCATCTGTCCAGGAAACTCCGACGGATCCTT 636
      |||
QY      181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
      |||
Db      637 CTGCCACTGTATGAAGTAGATGATTACGAGATGCCTTTAAAAACTCTGGGACTT 690
      |||

RESULT 10
AJ453647
LOCUS      696 bp mRNA linear EST 22-APR-2002
DEFINITION      AJ453647 riken1 Gallus gallus cdna clone 35b23rl, mRNA sequence.
ACCESSION      AJ453647
VERSION      AJ453647.1 GI:20263743
KEYWORDS      EST.
SOURCE      chicken.
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES             source
1..696
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="35b23rl"
/clone_lib="riken1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT      160 a 193 c 173 g 170 t
ORIGIN
Alignment Scores:
Pred. No.:      1.11e-116 Length:      696
Score:      977.00 Matches:      177
Percent Similarity: 93.94% Conservative: 9
Best Local Similarity: 89.39% Mismatches: 12
```

```

Query Match: 89.96% Indels: 0
DB: 9 Gaps: 0
US-09-966-880A-8 (1-198) x AJ453647 (1-696)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 97 ATGGACAGCCTCTTGATGAGGAAGAGCTTCTCTACAAATTTCAAGAACCTCGCTGG 156
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40
Db 157 GCCAAAGGCGTGTGAACCTACCTCTGTATGTTGTGAAGCGCGTGACAGTGCTACA 216
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db 217 TCATGCTCCTGGACTTGGATACCTGCGTAACAAGATGGTGGCTGCGAGGTCTC 276
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 277 TTCCCTACGCTATCATCTCAGCTTGGACCTGGACCCAGCGCGCTGCTACCGCATC 336
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 337 TTCACCTCTCGAGCGCTGATGACTGTGCGCGACATGTGGCTGACTTCTTCTGTCG 396
QY 101 AsnProAsnLeuSerLeuArgGluArgLeuHisArgAlaGlyValGlnIleAlaThr 140
Db 397 TACCCAAACTTGACCTCGCATTTTCTACTGCGCGCCTTACTTCTGTGAAGATCG 456
QY 121 AlaGluProGluGlyLeuArgLeuHisArgAlaGlyValGlnIleAlaThrMetThr 180
Db 457 GCTGAGCCTGAGGGCTGAGACGCTGACCGCGCTGGCGCCAAATCGCCATCATG 516
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 517 TTCAAGATTTCTTCTACTGCTGAACACGTTTGTGGAGAACAGGAAAGACATTCAA 576
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGlyIle 180
Db 577 GCCTGGGAAGGCTGCGATGAAACTCTGTCATCTGTCACAGAACTCTGGGACTT 636
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 637 CTGCACCTGTATGAAGTAGATGATTTACGAGATGCCCTTTTAAACTCTGGGACTT 696

RESULT 11
AJ450296 729 bp mRNA linear EST 19-APR-2002
LOCUS riken1 Gallus gallus cdna clone 24m22r1, mRNA sequence.
DEFINITION
ACCESSION AJ450296
VERSION AJ450296.1 GI:20217517
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Buerstedde,J.M.
1 (bases 1 to 729)
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
Location/Qualifiers
1. .729
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="24m22r1"
/clone_lib="riken1"
/cell_type="bursal lymphocyte"

FEATURES
source

```

```

/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 173 a 199 c 187 g 170 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 6,77e-108 Length: 729
Score: 910.00 Matches: 171
Percent Similarity: 92.75% Conservative: 8
Best Local Similarity: 88.60% Mismatches: 11
Query Match: 83.75% Indels: 3
DB: 9 Gaps: 0

```

US-09-966-880A-8 (1-198) x AJ450296 (1-729)

```

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 157 ATGGACAGCCTCTTGATGAGGAAGAGCTTCTCTACAAATTTCAAGAACCTCGCTGG 216
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40
Db 217 GCCAAAGGCGCTGTAACCTACCTCTGTATGTTGTGAAGCGCGTGACAGTGCTACA 276
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db 277 TCATGCTCCTGGACTTGGATACCTGCGTAACAAGATGGTGGCTGCGAGGTCTC 336
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 337 TTCCTACGCTATCATCTCAGCTTGGACCTGGACCCAGCGCGCTGCTACCGCATC 396
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 397 TTCACTCTCTGGAGCGCTCTGTATGACTGTGCGCGACATGTGCTGCTACTTCTCT 456
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 457 TACCCAAACTTGACCTCGCATTTTCACTGCGCGCCTCTACTTCTGTGAAGATCG 516
QY 121 AlaGluProGluGlyLeuArgLeuHisArgAlaGlyValGlnIleAlaThrMetThr 140
Db 517 GCTGAGCCTGAGGGCTGAGACGCTGACCGCGCTGGCGCCAAATCGCCATCATG 576
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 577 TTCAAGATTTCTTCTACTGCTGGAACACGTTTGTGGAGAACAGGAAAGACATTCAA 636
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGlyIle 180
Db 637 GCCTGGGAAGGCTGCGATGAAACTCTGTCATCTGTCACAGAACTCTGGGACTTCC 696
QY 180 eu-LeuProLeuTyrGluValAspAspLeuArg 190
Db 697 TTCCTGCCACTGTATGAAGTAGATGATTACGA 729

```

RESULT 12

AJ450317

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 688)

Buerstedde,J.M.

Gallus gallus bursal lymphocyte EST

Unpublished (2002)

Contact: Buerstedde JM

Cellular Immunology


```

Db      421 TTCTGTCGGGTAACACAGACAACCTCGCGCATCTTTGGCCCTGTGTACGAAGTCGATGA 480
Qy      188 pLeuArgAspAlaPheArgThrLeuGly 197
          |||||
Db      481 CTTGCGAGATGTCATTCGTATGTGGGA 508

RESULT 14
LOCUS   BF238155
DEFINITION 601811880F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054915 5',
mRNA sequence.
ACCESSION BF238155
VERSION   BF238155.1 GI:11152074
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 541)
           NIH-MGC http://mgi.nci.nih.gov/.
           National Institutes of Health, Mammalian Gene Collection (MGC)
           Unpublished (1999)
           Contact: Robert Strausberg, Ph.D.
           Email: cgabbs@mail.nih.gov
           Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/ILLNL at:
           http://image.llnl.gov
           Plate: LLCM895 row: p column: 20
           High quality sequence stop: 541.
           Location/Qualifiers
             1..541
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4054915"
               /clone_lib="NIH_MGC_48"
               /tissue_type="primary B-cells from tonsils (cell line)"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
               Site_2: EcoRI; cDNA made by oligo-dT priming.
               Directionally cloned into EcoRI/XhoI sites using the
               following 5' adaptor: GGCACGAG(G). Size-selected
               for average insert size 1.8kb. Library constructed by Ling
               Hong in the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies).
               Note: this is a NIH_MGC Library."
BASE COUNT 120 a 150 c 142 g 129 t
ORIGIN

Alignment Scores:
Pred. No.: 2,53e-101 Length: 541
Score: 858.00 Matches: 154
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.01% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BF238155 (1-541)

Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
          |||||
Db      78 ATGGACAGCCTCTTGATGAACCGGAGGAGTCTTCTTACCAATTCAGAAATGTCGCTGG 137

Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
          |||||
Db      138 GCTAAGGGTCGGGTGAGACCTACCTTGCTACGTAGTAGTAAGAGCGGTGACAGTCTACA 197

Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
          |||||

```

```

Db      198 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTGCTC 257
Qy      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
          |||||
Db      258 TTCTCCCGCTACATCTCGGACCTAGACCTGGCCGCTGCTACCGCTGACCTGG 317
Qy      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
          |||||
Db      318 TTACGCTCTGGAGGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTCTGGAGGG 377
Qy      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
          |||||
Db      378 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCCCTCTACTTCTGTGAGACGCGCAAG 437
Qy      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
          |||||
Db      438 GCTGAGCCGAGGGGCTGCGGGCTGCACCGCGGGGTGCAANTACCATCATGACC 497
Qy      141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsn 154
          |||||
Db      498 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAAC 539

RESULT 15
LOCUS   AJ449745
DEFINITION AJ449745 riken1 Gallus gallus cDNA clone 23a2r1, mRNA sequence.
ACCESSION AJ449745
VERSION   AJ449745.1 GI:20216966
KEYWORDS  EST.
SOURCE    chicken.
ORGANISM  Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 650)
           Buerstedde, J.M.
           Gallus gallus bursal lymphocyte EST
           Unpublished (2002)
           Contact: Buerstedde JM
           Cellular Immunology
           Heinrich-Pette-Institute
           Martinistr. 52, 20251 Hamburg, Germany
           Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
     source
       1..650
         /organism="Gallus gallus"
         /db_xref="taxon:9031"
         /clone="23a2r1"
         /clone_lib="riken1"
         /cell_type="bursal lymphocyte"
         /dev_stage="2-3 weeks old"
         /note="CB inbred strain"
BASE COUNT 150 a 182 c 167 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 3.82e-100 Length: 650
Score: 850.00 Matches: 156
Percent Similarity: 93.22% Conservative: 9
Best Local Similarity: 88.14% Mismatches: 12
Query Match: 78.27% Indels: 1
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AJ449745 (1-650)

```

```

Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
          |||||
Db      120 ATGGACAGCCTCTTGATGAACCGGAGGAGTCTTCTTACCAATTCAGAAATGTCGCTGG 179

Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
          |||||
Db      180 GCCAAAGCGCGTCGTGAACCTACCTCTGTTATCTTGTGAAGCCCGCTGACAGTGTACA 239

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model
Run on: June 19, 2003, 17:42:06 ; Search time 51 Seconds
(without alignments)
1190.627 Million cell updates/sec

Title: US-09-966-880A-8
Perfect score: 1086
Sequence: 1 MDSLMMRRKFLYQFKVNRW.....ILLPLYEVDDLRAFTLGL 198

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USFto.spool/US09966880/runat_14062003_175646_16331/app_query.fasta_1.391
-DB-Issued_Patents.NA -OFT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPLC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 @Cgn_1_1_40 @runat_14062003_175646_16331 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5b_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	369.5	34.0	610	1	US-08-816-241-2
2	369.5	34.0	610	3	US-09-128-395-2
3	271	25.0	891	1	US-08-687-895-2
4	271	25.0	891	2	US-09-040-482-2
5	210	19.3	879	1	US-08-158-682A-3
6	201	18.5	879	1	US-08-158-682A-1
7	201	18.5	879	1	US-08-015-203-1
8	87.5	8.1	1806	3	US-09-068-655-10
9	87.5	8.1	4713	3	US-09-068-655-3
10	80	7.4	7077	4	US-09-221-017B-123
11	77.5	7.1	1209	6	5352575-4
12	75.5	7.0	3043	4	US-09-049-698-16

C	13	75.5	7.0	3181	4	US-09-049-698-18	Sequence 18, Appl
	14	73.5	6.8	2629	1	US-08-200-807-1	Sequence 1, Appli
	15	73.5	6.8	2629	1	US-08-488-305A-1	Sequence 1, Appli
C	16	72.5	6.7	9179	4	US-09-453-702B-162	Sequence 162, App
	17	72.5	6.7	13188	4	US-08-961-527-70	Sequence 70, Appl
C	18	72	6.6	1672	4	US-09-495-066-1	Sequence 1, Appli
	19	72	6.6	4201	1	US-08-080-255-4	Sequence 4, Appli
	20	72	6.6	4201	3	US-08-465-713-4	Sequence 4, Appli
	21	72	6.6	4201	5	PCT-US93-05857-4	Sequence 4, Appli
	22	72	6.6	11907	4	US-08-061-376-4	Sequence 4, Appli
	23	72	6.6	14255	1	US-08-320-559-1	Sequence 1, Appli
	24	72	6.6	14255	1	US-08-327-392-1	Sequence 1, Appli
	25	72	6.6	14255	1	US-08-306-691B-55	Sequence 55, Appl
	26	72	6.6	14255	3	US-08-545-860D-1	Sequence 1, Appli
	27	72	6.6	14255	5	PCT-US94-04496-1	Sequence 1, Appli
	28	71	6.5	2534	5	PCT-US93-06251-10	Sequence 10, Appl
C	29	71	6.5	8220	2	US-08-568-459A-11	Sequence 11, Appl
	30	71	6.5	8220	2	US-08-487-826B-11	Sequence 11, Appl
	31	71	6.5	8220	4	US-09-210-288-11	Sequence 11, Appl
	32	71	6.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
C	33	70.5	6.5	704	4	US-08-998-416-977	Sequence 977, App
C	34	70	6.4	2156	2	US-08-899-514-1	Sequence 1, Appli
	35	70	6.4	2254	4	US-08-552-369-1	Sequence 1, Appli
	36	70	6.4	7616	5	PCT-US94-01149-54	Sequence 54, Appl
	37	69.5	6.4	1724	4	US-09-385-259-1	Sequence 1, Appli
	38	69.5	6.4	1724	4	US-09-645-370-1	Sequence 1, Appli
	39	69.5	6.4	6496	4	US-09-221-017B-543	Sequence 543, App
C	40	69	6.4	1035	4	US-08-191-160-4	Sequence 4, Appli
C	41	69	6.4	1770	4	US-08-191-160-3	Sequence 3, Appli
	42	69	6.4	7003	2	US-08-378-548-24	Sequence 24, Appl
C	43	69	6.4	8224	6	5180808-1	Patent No. 5180808
C	44	69	6.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C	45	69	6.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-816-241-2
; Sequence 2, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646833
US-08-816-241-2

Alignment Scores:
Pred. No.: 6,02e-38 Length: 610
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 4 Gaps: 4

US-09-966-880A-8 (1-198) x US-08-816-241-2 (1-610)
QY 11 PheLeuTyrglnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
Db 49 TTCTACTTCCAAATTTAAACACCTATGGGAAGCCACGATCGGAACAACTTGCTGTGC 108
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
Db 109 TTCACCGTGGAGGTATAAAGCCGCTCAGTTGCTCTCTGGAAGACG-----GGGTC 162
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db 163 TTCGGAACACGAGTGTATCTGAGACCCATTGTGCGAAGAGGTCTTCTCTCTGTGG 222
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrPheThrSerTyr 84
Db 223 TTCTGCGAGCAGTACTGTCTCTACACAAAGTACCAGTCACTTGGTACATCTGTG 282
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db 283 AGCCCTTGCCAGAGTGTGCGAGGGAGTGGCCGAGTTCCTGCGCAGGACGACGAGT 342
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrrPheCysGluAspArgLysAlaGluProGlu 124
Db 343 AATCTCACCATTCTACCGCCGCTCTACTACTTCCAG---TATCCATGTTACGAGG 399
QY 125 GlyLeuArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db 400 GGGCTCCGAGCCTGAGTCAGGAAGGGTGGCTGTGGAGATCATGGACATATGAAGATT 459
QY 145 PheTyrrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db 460 AATATTGTTGGAAACATTGTGTACAATGATAATGAGCCATTCAGGCTTTGGAAGGA 519
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 520 TTAANAACCACTTTCGACTTCTGAANAAGAGGCTACGGAGAGTCTC 567

RESULT 2
US-09-128-395-2
; Sequence 2, Application US/09128395
; Patent No. 6087108
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646833
US-09-128-395-2

Alignment Scores:
Pred. No.: 6,02e-38 Length: 610
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 4 Gaps: 4

US-09-966-880A-8 (1-198) x US-09-128-395-2 (1-610)
QY 11 PheLeuTyrglnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
Db 49 TTCTACTTCCAAATTTAAACACCTATGGGAAGCCACGATCGGAACAACTTGCTGTGC 108
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
Db 109 TTCACCGTGGAGGTATAAAGCCGCTCAGTTGCTCTCTGGAAGACG-----GGGTC 162
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db 163 TTCGGAACACGAGTGTATCTGAGACCCATTGTGCGAAGAGGTCTTCTCTCTGTGG 222
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrPheThrSerTyr 84
Db 223 TTCTGCGAGCAGTACTGTCTCTACACAAAGTACCAGTCACTTGGTACATCTGTG 282
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db 283 AGCCCTTGCCAGAGTGTGCGAGGGAGTGGCCGAGTTCCTGCGCAGGACGACGAGT 342
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrrPheCysGluAspArgLysAlaGluProGlu 124
Db 343 AATCTCACCATTCTACCGCCGCTCTACTACTTCCAG---TATCCATGTTACGAGG 399
QY 125 GlyLeuArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db 400 GGGCTCCGAGCCTGAGTACGGAAGGGTGGCTGTGGAGATCATGGACATATGAAGATT 459
QY 145 PheTyrrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db 460 AATATTGTTGGAAACATTGTGTACAATGATAATGAGCCATTCAGGCTTTGGAAGGA 519
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180

```


Db 520 TTAATAAACCAACTTTTCGACTTCTGAAAGAGAGGCTACGGGAGAGTCTC 567

RESULT 3

US-08-687-895-2
; Sequence 2, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0109 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 891 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: MUSCNOT1
CLONE: 57953

US-08-687-895-2

Alignment Scores:

Pred. No.: 3.73e-25 Length: 891

Score: 271.00 Matches: 63

Percent Similarity: 50.00% Conservative: 25

Best Local Similarity: 35.80% Mismatches: 76

Query Match: 24.95% Indels: 12

DB: 1 Gaps: 5

US-09-966-880A-8 (1-198) x US-08-687-895-2 (1-891)

QY 11 PheLeuTyrGlnPheLysValArgTrrAlaLysGlyArgGluThrTyrLeuCys 30
Db 306 TTTAATTCAGTTCGGGAATGGAGTGGAGTTCGGGAGGAGCAACACCTTCCTCTGC 365
QY 31 TyrValValLysArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArg 50
Db 366 TATGTGTTGAGCAGCAGGCGGAGGGGCGGAGTCCAGGCATCTCGGGGATACCTAGAG 425
QY 51 AsnLysAsn--GlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAsp 69
Db 426 GATGAGCATCGGCTGCCCATCGAGGAAGAGCTTTCTTCAACACCATCTGCCA---GCC 482
QY 70 LeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAsp 89
Db 483 TTCGACCCAGCCCTCGGGTACAAATGTCCTGTCTCCAGCCCTGTGAGCG 542
QY 90 CysAlaAlaHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePhe 109

Db 543 TGTGCTGACCGCATTAATAAACCTTAGCAAGCAACCAAGACCTGCTGCTCATTTCTG 602
QY 110 ThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu-----GlyLeu 126
Db 603 GTGGGTGCGACTCTTCATGTGGGAG-----GAGCCGGAGATCCAGGCTGCTCTG 650
QY 127 ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr 146
Db 651 AAGAAGCTGAAGGAGGCTGGCTCTAACTCCGATCATGAAGCCCGCAGGACTTCGAATAT 710
QY 147 CysTrpAsnThrPheValGluAsnHisGlu-----ArgThrPheLysAlaTrpGlu 163
Db 711 GTCTGGCAGAATTTCTGGAGCAAGAAGGGTGAATCCAAGGCCTTTCARCCCTGGGAG 770
QY 164 GlyLeuHisGluAsnSerValArgLeuSerArgGluLeuArgIle 179
Db 771 GACATTCAGGAGAACTTCCTATATACAGAGGAGAAAGTTGGCAGACATC 818

RESULT 4

US-09-040-482-2

; Sequence 2, Application US/09040482

; Patent No. 5916556

; GENERAL INFORMATION:

; APPLICANT: Au-Young, Janice

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Hillman, Jennifer L.

; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: U.S.

; ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

FILING DATE: US/09/040,482

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/687,895

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0109 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 891 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

IMMEDIATE SOURCE:

LIBRARY: MUSCNOT1

CLONE: 57953

US-09-040-482-2

Alignment Scores:

Pred. No.: 3.73e-25 Length: 891

Score: 271.00 Matches: 63

Percent Similarity: 50.00% Conservative: 25

Best Local Similarity: 35.80% Mismatches: 76

Query Match: 24.95% Indels: 12

DB: 2 Gaps: 5

```
US-09-966-880A-8 (1-198) x US-09-040-482-2 (1-891)
Qy 11 PheLeuTyrGlnPheLysAsnValArgTirpAlaLysGlyArgGluThrTyrLeuCys 30
Db 306 TTTAAATCCAGTTCGGATGTGGATACAGTCCGGGAGGACACAGACCTTCTCTGC 365
Qy 31 TyrValValLysArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArg 50
Db 366 TATGTGTTGAAGCACACAGGGGAGGGGGCCAAAGTCAGGCATCTCGGGGATACCTAGAG 425
Qy 51 AsnLysAsn---GlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAsp 69
Db 426 GATGAGCATCGGCTGCCATGACGAGGAAGCTTTCTCAACACCATCTCTGCCA---GCC 482
Qy 70 LeuAspProGlyArgCysTyrArgValThrTirpPheThrSerTrpSerProCysTyrAsp 89
Db 483 TTCGACCCAGCCCTGCGGTACATGTCACTGTGTGTCTCCTCAGCCCTGTGCAGCG 542
Qy 90 CysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePhe 109
Db 543 TGTGCTGACCGCATTAACAAACCTTAGCAAGACCAAGAACCTGCGTCTCTCATCTG 602
Qy 110 ThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu-----GlyLeu 126
Db 603 GTGGTGCATCTTCATGTGGAG-----GAGCGGAGATCCAGGCTGCTCTG 650
Qy 127 ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr 146
Db 651 AAGAAGCTGAAGGAGGCTGCTGTAACTCGCATCATGAAGCCCAAGGCTTCGATAT 710
Qy 147 CysTrpAsnThrPheValGluAsnHisGlu-----ArgThrPheLysAlaTrpGlu 163
Db 711 GTCTGGCAGAAATTTGTGGAGCAAGAGGGTGAATCCCAAGGCTTTTCARCCCTGGGAG 770
Qy 164 GlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIle 179
Db 771 GACATTCAGGAGAACTCTTACTACTAGCAGGAGAAAGTTGGCAGACATC 818

RESULT 5
US-08-158-682A-3
; Sequence 3, Application US/08158682A
; Patent No. 5434058
; GENERAL INFORMATION:
; APPLICANT: Davidson, Nicholas O.
; TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: 321 No. 5434058th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,682A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coolley, Ronald B.
; REGISTRATION NUMBER: 27,187
; REFERENCE/DOCKET NUMBER: ARCD:085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 23..730
US-08-158-682A-3
Alignment Scores:
Pred. No.: 2,15e-17 Length: 879
Score: 210.00 Matches: 48
Percent Similarity: 56.30% Conservative: 28
Best Local Similarity: 35.56% Mismatches: 43
Query Match: 19.34% Indels: 16
DB: 1 Gaps: 5
US-09-966-880A-8 (1-198) x US-08-158-682A-3 (1-879)
Qy 24 ArgArgGluThrTyrLeuCysTyrValValLys-----ArgArgAspSerAla 39
Db 119 CGTAAAGAGGCGCTGCTGCTCTACGAAATCAAGTGGGCATGAGCCGGAAGATCTGGCGA 178
Qy 40 ThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCys---HisValGlu 58
Db 179 ACGTCA-----GGCAAAAACACACCACCAATCACGTGCA 211
Qy 59 LeuLeuPheLeu---ArgTyrIleSerAspTrpAspLeuAspProGlyVArgCysTyrArg 77
Db 212 GTTAATTATATAAAAAATTTACGTCAGAAAGATTTTACCCATCATCAGCTGCACC 271
Qy 78 ValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPhe 97
Db 272 ATCAGCTGGTCTTGTCTCTGGAGTCCCTGCTGGGAATGCTCCAGGCTATTAGAGAGTTT 331
Qy 98 LeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu 117
Db 332 CTGAGTCGGCACCCTGGTGTGACTGTACTGTACTGTACGTAGCTCGGCTTTTGGCAGCATG 391
Qy 118 AspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAla 137
Db 392 GATCAACAA---AATCGCAAGGTCTCAGGACCTTGTTAACAGTGGAGTAATTTCAG 448
Qy 138 IleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheVal 152
Db 449 ATTATGAGAGCATCAGAGTATTATCACTGCTGGAGGAATTTTGTTC 493

RESULT 6
US-08-158-682A-1
; Sequence 1, Application US/08158682A
; Patent No. 5434058
; GENERAL INFORMATION:
; APPLICANT: Davidson, Nicholas O.
; TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: 321 No. 5434058th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,682A
; FILING DATE:
; CLASSIFICATION: 435
```


CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 PAGE MILL ROAD
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows
 SOFTWARE: FastSeq for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/221,017B
 FILING DATE: 23-DEC-1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1182
 FILING DATE: 31-DEC-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1546
 FILING DATE: 30-JAN-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP2911
 FILING DATE: 09-APR-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU98/01023
 FILING DATE: 10-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Monroy, Gladys H
 REGISTRATION NUMBER: 32,430
 REFERENCE/DOCKET NUMBER: 27340-20021.00
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141

INFORMATION FOR SEQ ID NO: 123:

SEQUENCE CHARACTERISTICS:
 LENGTH: 7077 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: UNKNOWN
 ORIGINAL SOURCE:
 ORGANISM: PORYPHYROMONAS GINGIVALIS
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1...7077
 US-09-221-017B-123

Alignment Scores:
 Pred. No.: 17.5 Length: 7077
 Score: 80.00 Matches: 22
 Percent Similarity: 47.62% Conservative: 8
 Best Local Similarity: 34.92% Mismatches: 17
 Query Match: 7.37% Indels: 16
 DB: 4 Gaps: 3

US-09-966-880A-8 (1-198) x US-09-221-017B-123 (1-7077)

QY 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGly 23
 DB 3900 TTGTATATAATCGAAGACGTTTATATATAGATCTCAGCCGTTTCGATGGGAAAGAC 3959
 QY 24 ArgArgGluThrTyrLeuCysTyrValValLys-----ArgArgAspSerAlaThr 40
 DB 3960 GCCGAAGAACITTTCTTGAAACCATAAAAACCAAGGCCGCCGCGAAGTTCCTGT--- 4016
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
 DB 4017 -----CGAACTCCGAGCGGGTGC-----CTTGCT 4040

QY 61 PheLeuArg 63
 DB 4041 TTTGTGCT 4049
 RESULT 11
 5352575-4
 ;Patent No. 5352575
 ; APPLICANT: PETROVSKIS, ERIK A.;POST, LEONARD E.;TIMMINS, JAMES G.
 ; TITLE OF INVENTION: PSEUDORABIES VIRUS PROTEIN
 ; NUMBER OF SEQUENCES: 12
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/513,282
 ; FILING DATE: 20-APR-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 100,817
 ; FILING DATE: 29-JUN-1987
 ; APPLICATION NUMBER: 886,260
 ; FILING DATE: 16-JUL-1986
 ; APPLICATION NUMBER: 784,787
 ; FILING DATE: 04-OCT-1985
 ; APPLICATION NUMBER: 801,799
 ; FILING DATE: 26-NOV-1985
 ; APPLICATION NUMBER: 844,113
 ; FILING DATE: 26-MAR-1986
 ; SEQ ID NO: 4:
 ; LENGTH: 1209
 5352575-4

Alignment Scores:

Pred. No.: 2.6 Length: 1209
 Score: 77.50 Matches: 42
 Percent Similarity: 34.91% Conservative: 17
 Best Local Similarity: 24.85% Mismatches: 73
 Query Match: 7.14% Indels: 37
 DB: 6 Gaps: 6

US-09-966-880A-8 (1-198) x 5352575-4 (1-1209)

QY 24 ArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSer 43
 DB 262 CGCGGCGCCACGTACCGCGCCACGTG-----288
 QY 44 LeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeuPheLeuArg 63
 DB 289 -----GCCTGGTACCGCATCGCGGCGGGTGCACACCTGCTGTACTTTATCGAG 339
 QY 64 TyrIleSerAspTrpAspLeuAspPro-----GlyArgCysTyrArgValThr 79
 DB 340 TAGGCC-----GACTCGACCCCGCAGGAGGTCTTTGGGCGCTGCGGCGCGCAC 390
 QY 80 TrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArg 99
 DB 391 ACGCCGATGTGGTGACCCCGCGCGGACTACATGTCCCGCGGAGGAGAGCTG--- 447
 QY 100 GlyAsnProAsnLeuSerLeuArgIlePhePheAlaArgLeuTyrPheCysGluAspArg 119
 DB 448 -----GGGCTGCTCATGTTGGCCCCCGGGGG 474
 QY 120 LysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGluValGlnIleAla----- 137
 DB 475 TTCAACGAGGCGCAGTACCGCGCGCTGTGTCGCTCGCAGCGGTGAACATCTCACCAC 534
 QY 138 ---IleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGlu 156
 DB 535 TTGATGGTGGCGCTCCCGCGGGGCAAGAGTGCCTTCGCCCGCGTGGACGACAC--- 591
 QY 157 ArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeu 176
 DB 592 CGCACGTACAGTTCGGCGCGTGTGGAGCGACGACAGCTCAAGCGGGCGGTGACGTG 651
 QY 177 ArgArgIleLeuLeuProLeuTyrGlu 185

Db 652 ATGCCATTCCTGACGCCGCTTCTACCAG 678

RESULT 12

US-09-049-698-16/c

; Sequence 16, Application US/09049698

; Patent No. 6368792

; GENERAL INFORMATION:

; APPLICANT: BILLING-MEDEL, PATRICIA A.

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: HAYDEN, MARK

; APPLICANT: KLASS, MICHAEL R.

; APPLICANT: ROBERTS-RAPP, LISA

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STROUPE, STEPHEN D.

; TITLE OF INVENTION: REAGENTS AND METHODS FOR THE

; TITLE OF INVENTION: USEFUL FOR DETECTING DISEASES OF THE GASTROINTESTINAL

; TITLE OF INVENTION: TRACT

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/049,698

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/828,856

; FILING DATE: 31-MAR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6068. US.P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3043 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-049-698-16

Alignment Scores:

Pred. No.: 18 6 Length: 3043

Score: 75.50 Matches: 32

Percent Similarity: 43.09% Conservative: 21

Best Local Similarity: 26.02% Mismatches: 49

Query Match: 6.95% Indels: 21

DB: 4 Gaps: 5

US-09-966-880A-8 (1-198) x US-09-049-698-16 (1-3043)

QY 58 GluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys----- 75

Db 376 GAACCTGGTGGTGTATGTTCTCTACCTGGAG-----TGTAGGTGG 332

QY 76 -----TyrArgValThrPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 331 TCAACTATAACATCAGCATGTTATGTTTTCANGCTTTTGGCCCTTTTGTACTGAGGATT 272

QY 94 ValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeu 113

Db 271 TTCCTTCCAAATCTCAGG-----AATTATATAGATACATTTTTCAAAAAAATCT 221

QY 114 TyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGly 133

Db 220 TTTTCTGT-----GGCTTCAACACAGGTAGCTAGAGCTGT 185

QY 134 ValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGlu 153

Db 184 AGTCACCATATCCTCTATTGTTCAATATTATTTTTCATCTTTCGCACACATAGGATCTAT 125

QY 154 AsnHisGluArgThrPheLysAla---TrpGluGlyLeuHisGluAsnSerValArgLeu 172

Db 124 AACATGACATATCTTCAAGCCATATTATTTCAGCTTATATGAAGGAGTATTTGACATG 65

QY 173 SerArgGln 175

Db 64 GTGCAGCAG 56

RESULT 13

US-09-049-698-18/c

; Sequence 18, Application US/09049698

; Patent No. 6368792

; GENERAL INFORMATION:

; APPLICANT: BILLING-MEDEL, PATRICIA A.

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: HAYDEN, MARK

; APPLICANT: KLASS, MICHAEL R.

; APPLICANT: ROBERTS-RAPP, LISA

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STROUPE, STEPHEN D.

; TITLE OF INVENTION: REAGENTS AND METHODS FOR THE

; TITLE OF INVENTION: USEFUL FOR DETECTING DISEASES OF THE GASTROINTESTINAL

; TITLE OF INVENTION: TRACT

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/049,698

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/828,856

; FILING DATE: 31-MAR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6068. US.P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3181 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-049-698-18

```
Alignment Scores:
Pred. No.: 19.8 Length: 3181
Score: 75.50 Matches: 32
Percent Similarity: 43.09% Conservative: 21
Best Local Similarity: 26.02% Mismatches: 49
Query Match: 6.95% Indels: 21
DB: 4 Gaps: 5

US-09-966-880A-8 (1-198) x US-09-049-698-18 (1-3181)
QY 58 GluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys----- 75
Db 387 GAACTGCTGGTGTTCATCTCTACCTGGAG-----TGTAGGTGG 343
QY 76 -----TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93
Db 342 TGCAACTATAACATCAGCATGTTTATGCTTTTCATGTTTGGCCTTTTGTACGAGGATT 283
QY 94 ValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeu 113
Db 282 TTCTTCCCAATTCTCAGG-----AATTAATATAGATACATTTTGGAAAAAATCT 232
QY 114 TyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGly 133
Db 231 TTTTCTGT-----GGCTTCAACAGGCTACGTAGAAGCTGT 196
QY 134 ValGluIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGlu 153
Db 195 AGTCACCATATCTCTATTGTTCAATTATTTTTCATCTCTGCGCACACTAGATCTAT 136
QY 154 AsnHisGluArgThrPheLysAla---TrpGluGlyLeuHisGluAsnSerValArgLeu 172
Db 135 ACAAATGACAATATCTTCAAGGCCATTATTATTACAGCTTAATGAAGGAAGTATTGACTG 76
QY 173 SerArgGln 175
Db 75 GTGCAGCAG 67

RESULT 14
US-08-200-807-1
; Sequence 1, Application US/08200807
; Patent No. 5573939
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof, Eriksson, Ulf
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies which
; bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5573939
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,807
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/883,539
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5573939man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 280
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; US-08-200-807-1
Alignment Scores:
Pred. No.: 26.8 Length: 2629
Score: 73.50 Matches: 29
Percent Similarity: 37.93% Conservative: 26
Best Local Similarity: 20.00% Mismatches: 43
Query Match: 6.77% Indels: 47
DB: 1 Gaps: 8

US-09-966-880A-8 (1-198) x US-08-200-807-1 (1-2629)
QY 57 ValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyr 76
Db 782 GTCAAAATTAATCTGTTCAAGTTTCTTCTTCATGGAGTCTT----- 823
QY 77 ArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg----- 92
Db 824 -----TGGGAGGCCAATTACATGATGTTGTTTGAATCCAATGAACCATGGGGTT 874
QY 93 -----HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThr 110
Db 875 TGCCTTCATATTGCTGACAAAAAAGA----- 901
QY 111 AlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHis 130
Db 902 ---AAAAAGTATATATCAATAATAACAGGACCTCTCTTTTAACCTC-----TTTCAT 952
QY 131 ArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr----- 146
Db 953 CAC-----ATCAATACCTATGAAGACCATGAGTTTCTGATTTGCGAT 994
QY 147 -----CysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAla----- 161
Db 995 CTCGTGCTGGAAGGATTGAAATTTGTTTATATATTATTATTATTAGCCCAATTTACGT 1054
QY 162 -----TrpGluGlyLeuHisGluAsnSerValArgLeuSerArg---GlnLeuArgArg 178
Db 1055 GAGAACTGGGAAGAGGTGAAAAAAATGCCAGAAAGGCTCTCAGCCTGAAGTTAGGAGA 1114
QY 179 IleLeuLeuProLeu 183
Db 1115 TAGCTACTTCTCTTG 1129

RESULT 15
US-08-488-305A-1
; Sequence 1, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof, Eriksson, Ulf; Peterson, Per A.
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies which
; bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,305A
FILING DATE: 7-JUNE-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kohli, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5280.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2629 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
US-08-488-305A-1

Alignment Scores:
Pred. No.: 26.8 Length: 2629
Score: 73.50 Matches: 29
Percent Similarity: 37.93% Conservative: 26
Best Local Similarity: 20.00% Mismatches: 43
Query Match: 6.77% Indels: 47
DB: 1 Gaps: 8

US-09-966-880A-8 (1-198) x US-08-488-305A-1 (1-2629)

QY	57	ValGlulLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyr	76
Db	782	GTCAAAATTAATCTGTTCAAGTTCTTTCTTCATGGAGTCTT	823
QY	77	ArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg	92
Db	824	-----TCGGGAGCCAATTACATGGATTGTTTGAATCCAAATGAACCATGGGGTT	874
QY	93	-----HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThr	110
Db	875	TGGCTTCATATGCTGACAAAAAAGA	901
QY	111	AlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHis	130
Db	902	---AAAAAGTATATCAATAATAAATACAGGACCTCTCCTTTTAACCTC-----TTTCAT	952
QY	131	ArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr	146
Db	953	CAC-----ATCAATACCTATGAGACCATGAGTTTCTGATTGTGGAT	994
QY	147	-----CysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAla	161
Db	995	CTCTGTCTGGAAGGATTGTAATTGTTTATTAATATTATATTAGCCAATTACGT	1054
QY	162	-----TrpGluGlyLeuHisGluAsnSerValArgLeuSerArg---GlnLeuArgArg	178
Db	1055	GAGAACTGGGAGAGGTGAAAAAATGCCAGAAAGGCTCTCTCAGCCTGAAGTTAGGAGA	1114
QY	179	IleLeuLeuProLeu	183
Db	1115	TACGTACTTCCTTTG	1129

Search completed: June 19, 2003, 19:04:05
Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 19, 2003, 17:54:11 ; Search time 156 Seconds
(without alignments)
1862.504 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDLLMNRKFLYQKNRW.....ILLPLYVDLDRFRTGL 198

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09966880/runat_14062003_175646_16343/app_query.fasta_1.391
-DB=PublishedApplications_NA -Qfmt=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NAPLEN=2000000000 -USER=US09966880@cgn_1.1.57_runat_14062003_175646_16343
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_1/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	2818	9 US-09-966-880A-7	Sequence 7, Appli
2	1008	92.8	2440	9 US-09-966-880A-1	Sequence 1, Appli
3	644.5	59.3	6364	9 US-09-966-880A-10	Sequence 10, Appl
4	644.5	59.3	11204	9 US-09-966-880A-35	Sequence 35, Appl

5	505	46.5	271	9 US-09-966-880A-13	Sequence 13, Appl
6	390	35.9	1534	10 US-09-729-674-173	Sequence 173, App
7	369.5	34.0	987	10 US-09-925-300-699	Sequence 699, App
8	369.5	34.0	1143	9 US-10-098-841-268	Sequence 268, App
9	363.5	33.5	1348	10 US-09-880-107-3282	Sequence 3282, Ap
10	308.5	28.4	2151	10 US-09-822-830A-359	Sequence 359, App
11	260	23.9	148	9 US-09-966-880A-12	Sequence 12, Appl
12	257	23.7	371	9 US-09-918-995-30237	Sequence 30237, A
13	238.5	22.0	1404	9 US-10-198-846-13472	Sequence 13472, A
14	218.5	20.1	879	9 US-10-157-031-13	Sequence 13, Appl
15	212	19.5	116	9 US-09-966-880A-14	Sequence 14, Appl
16	206	19.0	390	9 US-09-796-692-9051	Sequence 9051, Ap
17	206	19.0	390	9 US-10-040-862-9051	Sequence 9051, Ap
18	175	16.1	556	9 US-09-796-692-5061	Sequence 5061, Ap
19	175	16.1	556	9 US-10-040-862-5061	Sequence 5061, Ap
20	166.5	15.3	476	9 US-09-918-995-32385	Sequence 32385, A
21	164	15.1	553	9 US-09-796-692-4504	Sequence 4504, Ap
22	164	15.1	553	9 US-10-040-862-4504	Sequence 4504, Ap
23	162.5	15.0	572	10 US-09-864-761-9416	Sequence 9416, Ap
24	157.5	14.5	328	9 US-09-796-692-5737	Sequence 5737, Ap
25	157.5	14.5	328	9 US-10-040-862-5737	Sequence 5737, Ap
26	122	11.2	505	9 US-09-796-692-4443	Sequence 4443, Ap
27	122	11.2	505	9 US-10-040-862-4443	Sequence 4443, Ap
28	114	10.5	261	9 US-10-015-219-763	Sequence 763, App
29	114	10.5	261	10 US-09-777-564-763	Sequence 763, App
30	112.5	10.4	466	9 US-09-796-692-4007	Sequence 4007, Ap
31	112.5	10.4	466	9 US-10-040-862-4007	Sequence 4007, Ap
32	101	9.3	559	10 US-09-864-761-9458	Sequence 9458, Ap
33	99.5	9.2	473	9 US-10-046-935-1443	Sequence 1443, Ap
34	99.5	9.2	473	9 US-09-878-178-1443	Sequence 1443, Ap
35	99.5	9.2	473	9 US-10-146-502-1443	Sequence 1443, Ap
36	97.5	9.0	254	10 US-09-867-701-3576	Sequence 3576, Ap
37	88	8.1	2172	9 US-09-966-880A-15	Sequence 15, Appl
38	84	7.7	154	10 US-09-864-761-25835	Sequence 25835, A
39	84	7.7	997	10 US-09-864-761-9278	Sequence 9278, Ap
40	82.5	7.6	415	10 US-09-764-877-3699	Sequence 3699, Ap
41	81.5	7.5	14708	9 US-10-239-676-221	Sequence 221, App
42	80	7.4	555	10 US-09-864-761-9282	Sequence 9282, Ap
43	78	7.2	342	10 US-09-783-590-5098	Sequence 5098, Ap
44	77.5	7.1	1437	9 US-10-166-087-49	Sequence 49, Appl
45	77.5	7.1	1992	9 US-10-080-960-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)

```

; FEATURE:
; NAME/KEY: 5' UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3' UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

```

Alignment Scores:	
Pred. No.:	6.66e-142
Score:	1086.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	9
Length:	2818
Matches:	198
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-966-880A-8 (1-198) x US-09-966-880A-7 (1-2818)

1	Met	Asp	Ser	Leu	Leu	Met	Asn	Arg	Arg	Gly	Ser	Phe	Leu	Tyr	Gln	Phe	Lys	Asn	Val	Arg	Trp	20		
80	AT	GC	AG	AC	GC	CT	TT	GAT	GA	CC	GG	AG	GA	GT	TT	CT	T	TAC	CAA	AT	T	GC	CG	139
21	Ala	Lys	Gly	Arg	Arg	Glu	Thr	Tyr	Leu	Cys	Tyr	Val	Val	Lys	Arg	Arg	Asp	Ser	Ala	Thr	40			
140	GCT	AAG	GT	CG	CG	TG	TAC	CT	CT	G	T	GT	CT	GT	GA	AG	GG	CG	T	G	AC	GT	199	
41	Ser	Phe	Ser	Leu	Asp	Phe	Gly	Tyr	Leu	Arg	Asn	Lys	Asn	Gly	Cys	His	Val	Glu	Leu	Leu	60			
200	TC	CT	TT	T	CAC	T	GG	ACT	TT	G	GT	TAT	CT	T	CG	CA	TA	AGA	AC	GG	CT	GC	259	
61	Phe	Leu	Arg	Tyr	Ile	Ser	Asp	Trp	Asp	Leu	Asp	Pro	Gly	Arg	Cys	Tyr	Arg	Val	Thr	Trp	80			
260	TT	CC	T	CC	CG	CT	AC	AT	CT	CG	GA	CT	AG	AC	CT	GG	CG	CT	GC	T	AC	CG	319	
81	Phe	Thr	Ser	Trp	Ser	Pro	Cys	Tyr	Asp	Cys	Ala	Arg	His	Val	Ala	Asp	Phe	Leu	Arg	Gly	100			
320	TT	CAC	CT	CCT	CG	AG	CC	CT	GT	CT	AG	CT	GT	CG	CG	AC	AT	GT	G	CC	AC	TT	379	
101	Asn	Pro	Asn	Leu	Ser	Leu	Arg	Ile	Phe	Thr	Ala	Arg	Leu	Tyr	Phe	Cys	Glu	Asp	Arg	Lys	120			
380	AA	CC	CA	CA	CC	CT	CAG	T	CT	GA	GC	AT	CT	CA	CC	CG	CG	CT	CT	CT	CT	GA	439	
121	Ala	Glu	Pro	Glu	Gly	Leu	Arg	Arg	Leu	His	Arg	Ala	Gly	Val	Gln	Ile	Ala	Ile	Met	Thr	140			
440	GCT	AG	CC	CG	AG	GG	CT	CG	CG	CG	CT	CG	AC	CG	CG	GG	GT	G	CAA	AT	AG	CA	499	
141	Phe	Lys	Asp	Tyr	Phe	Tyr	Cys	Trp	Asn	Thr	Phe	Val	Glu	Asn	His	Glu	Arg	Thr	Phe	Lys	160			
500	TT	CA	AA	GA	T	T	T	T	T	ACT	GC	T	GG	A	T	ACT	T	T	T	T	T	T	559	
161	Ala	Trp	Glu	Gly	Leu	His	Glu	Asn	Ser	Val	Arg	Leu	Ser	Arg	Gln	Leu	Arg	Arg	Ile	Leu	180			
560	GC	CT	GG	GA	GG	CT	GC	AT	G	AA	AT	T	CAG	T	CG	T	CT	C	C	AG	AC	GT	619	
181	Leu	Pro	Leu	Tyr	Glu	Val	Asp	Asp	Leu	Arg	Asp	Ala	Phe	Arg	Thr	Leu	Gly	Leu	198					
620	TT	GC	CC	CT	GT	AT	G	GT	TT	G	AT	CT	T	AC	G	AC	GC	AT	T	T	CG	ACT	673	

RESULT 2

```

RECORDED-96-880A-1
: Sequence 1, Application US/09966880A
: Patent No. US20020164743A1
: GENERAL INFORMATION:
: APPLICANT: Honjo, Tasuku
: APPLICANT: Muramatsu, Masamichi
: TITLE OF INVENTION: NOVEL CYTIDINE
: FILE REFERENCE: 06501-038001
: CURRENT APPLICATION NUMBER: US/09/9
: CURRENT FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: PCT/JPO00/
: PRIOR FILING DATE: 2000-03-28
: PRIOR APPLICATION NUMBER: JP 11-371
: PRIOR FILING DATE: 1999-12-27

```

Db 633 TTGCCCTGTACGAAGTCGATGCACTTGCAGATGCATTTCCGTATGTTGGGA 683

RESULT 3

US-09-966-880A-10

; Sequence 10, Application US/09966880A

; Patent No. US20020184743A1

; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku

; APPLICANT: Muramatsu, Masamichi

; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE

; FILE REFERENCE: 06501-088001

; CURRENT APPLICATION NUMBER: US/09/966,880A

; CURRENT FILING DATE: 2001-09-28

; PRIOR APPLICATION NUMBER: PCT/JP00/01918

; PRIOR FILING DATE: 2000-03-28

; PRIOR APPLICATION NUMBER: JP 11-3711382

; PRIOR FILING DATE: 1999-12-27

; PRIOR APPLICATION NUMBER: JP 11-178999

; PRIOR FILING DATE: 1999-06-24

; PRIOR APPLICATION NUMBER: JP 11-87192

; PRIOR FILING DATE: 1999-03-29

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10

; LENGTH: 6564

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-966-880A-10

Alignment Scores:

Pred. No.: 3.11e-79 Length: 6564

Score: 644.50 Matches: 177

Percent Similarity: 24.02% Conservative: 0

Best Local Similarity: 24.02% Mismatches: 2

Query Match: 59.35% Indels: 560

DB: 9 Gaps: 2

US-09-966-880A-8 (1-198) x US-09-966-880A-10 (1-6564)

QY 3 SerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLys 22

Db 1063 AGCCCTTGATGAACGGAGGAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAG 1122

QY 23 GlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThrSerPhe 42

Db 1123 GGTGCGGTGAGACCTACTGCTACGTAGTAGTAAGAGCGGTGACAGTGTACATCCCTTT 1182

QY 43 SerLeuAspPheGlyTyrLeuArgAsnLys

Db 1183 TCACCTGAGCTTTGGTTATCTTCGCAATAA-GGTATCAATTAAGTCAGCTTTTGCAAGCAG 1241

QY 52

Db 1242 TTAAATGTCAACTGTGAGTGCTTTTAGAGCCACCTGTGATGGTATTACTTCCATCCTT 1301

QY 52

Db 1302 TTTTGGCATTTGTCTCTATCACATTCCTCAAAATCCTTTTTTTATTTCTTTTCCATG 1361

QY 52

Db 1362 TCCATGCCCATATTAGACATGGCCAAAATATGTGATTTAATTCCTCCCCAGTAATGC 1421

QY 52

Db 1422 TGGGACCCCTAATACCCTCTTCTTTCAGTGCCCAAGAACAACTGTCTCCAAACTGTTTA 1481

QY 52

Db 1482 CCAGCTTTCCCTCAGCATCTGAATTCGCTTTTGAGATTAATTAAGCTAAAGCATTTTATA 1541

QY 52

Db 1542 TGGGAGAAATATTATCAGCTTGTCCAGCAAAAATTTTAAATGTGAAAAACAAATTTGTGTC 1601

QY 52

Db 1602 TTAAGCATTTTGGAAAAATTAAGGAAGAAGAAATTTGGGAAAAAATTAACGGTGTTCAAATT 1661

QY 52

Db 1662 CTGTTTTCCAAATGATTCTTTTCCCTCTACTCACATGGTGTAGCCAGTAATACA 1721

QY 52

Db 1722 TTCAACATGGTGATCCCCAGAAAACCTCAGAGAAGCCTCGGCTGATGATTAAATTAATGA 1781

QY 52

Db 1782 TCCTTCGGGTACCCGAGAGAATTACATTTCCAAGAGACTTCCTCACCAAAATCCAGATGG 1841

QY 52

Db 1842 GTTTACATAAACTTCTGCCCATGGGTATCTCTCTCTTAACACGCTGTGACGCTCTGGG 1901

QY 52

Db 1902 CTTGTGGAATCTCAGGGAAGCATCCGTGGGTGGAAGGTCAATCGTCTGGCTCGTTGTTT 1961

QY 52

Db 1962 GATGGTTATATTACCATGCAATTTCTTTTGCCTACATTTGTATTGAATACATCCCAATCT 2021

QY 52

Db 2022 CCTTCTATTTCGGTGACATGACACATTTCTATTTCAGAAGGCTTTGATTTTATCAAGCACT 2081

QY 52

Db 2082 TTCAATTTACTTCTCATGGCAGTGCTTACTTCTTTACAATPACCATCTGTCTGCTTT 2141

QY 52

Db 2142 ACCAAAATCTATTTCCCTCTTTTCAGATCTCCCAATGTCTCATAAACTGTCTGCTCCT 2201

QY 52

Db 2202 CCACCTAGTGGTCCAGGTATATTTCACAAATGTTACATCAACAGGCACTTCTAGCCATTT 2261

QY 52

Db 2262 TCCTTCTCAAAAGGTGCAAAAAGCAACTTTCATAAACACAAAATTTAAATCTTTCGGTGAGTA 2321

QY 52

Db 2322 GTGTGATGTGCTTCTCTCCAACTCAGCGCACTTCGCTTCCTTCATTCCACAAAAACCCA 2381

QY 52

Db 2382 TAGCCTTCTTCACTCTGCGGACTAGTGTGCCAAGGGTTACAGCTTACCTACTGGTGT 2441

QY 52

Db 2442 GCTCTTTTGAGCAAGTTGCTTAGCCCTCTCTGTAACACAAAGACAATAGCTGCAAGCATCC 2501

QY 52

Db 2502 CCAAGATCATTTGCAGGAGACAATGACTAAGGTACCAGCGCGCAATAAAAGTCAGTGA 2561

QY 53

Db 2562 ATTTAGCGTGGTCTCTCTCTCTCTCCAGAGCGGTGCCACGTGGNAATGCTCTCTCCTC 2621

QY 63 ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThr 82

Db 2622 CGCTACATCTCGAGCTGGGACCTAGACCTGGCGCTGTACCGCGTCACTGTTCCACC 2681

QY 52 ----- 52
Db 7297 TTCAATTACTTCTCATGCGAGTGCCTATTACTTCTCTTACAATAACCCATCTCTCGCTTT 7356
QY 52 ----- 52
Db 7357 ACCAAAATCTATTTCCTCCCTTTTCAGATCCCTCCAAATGGTCTCTCATAAACTGTCTCTGCT 7416
QY 52 ----- 52
Db 7417 CCACCTAGTGGTCCAGGTATATTTCACAATGTTACATCAACAGGCACTTCTAGCCATT 7476
QY 52 ----- 52
Db 7477 TCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACAATAATAATCTTCGGTGAGGTA 7536
QY 52 ----- 52
Db 7537 GTGTGATGCTGCTTCTCCCAACTCAGCGCACTTGTCTTCCTCATTCACAAAACCCA 7596
QY 52 ----- 52
Db 7597 TAGCCTTCTTCACTCTGCAGGACTAGTGTCTGCCAAGGTTTCAGCTCTACCTACTTGGTCT 7656
QY 52 ----- 52
Db 7657 GCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAGGACAATAGTGTCAAGCATCC 7716
QY 52 ----- 52
Db 7717 CCAAGATCATTGCAGGAGACAATGACTAAGGCTACCAGCGCCAAATAAAGTCAGTGA 7776
QY 53 ----- AsnGlyCysHisValGluLeuLeuPheLeu 62
Db 7777 ATTTTAGCGTGTCTCTCTGTCTCTCCAGAACGCTGCCAGTGGAAATGTCTCTCTC 7836
QY 63 ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThr 82
Db 7837 CGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTACCTGCCGTCACTGGTTCACC 7896
QY 83 SerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnPro 102
Db 7897 TCCTGGAGCCCTGTAGACTGTGCCGACATGTGGCGACTTTCTCGAGGGAACCC 7956
QY 103 AsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGlu 122
Db 7957 AACCTCAGTCTGAGGATCTTACCGGGCCCTTACTTCTGTGAGGACCGCAAGGCTGAG 8016
QY 123 ProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLys 142
Db 8017 CCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACCTTCAA- 8075
QY 142 ----- 142
Db 8076 AGGTCCGAAGGGCCCTTCGGCGCAGGCGCAGTGCAGACGCCCGCATTCGGGATTCGATG 8135
QY 142 ----- 142
Db 8136 CGGAATGAATGAGTTAGTGGGAAGCTCGAGGGGAAGAGTGGCGGGGATTCGTTC 8195
QY 142 ----- 142
Db 8196 CCTCTGAGCCGAAATTAAGATTAGAACGACAGAGAAAAGTGAATGGCTCAGACACAAG 8255
QY 142 ----- 142
Db 8256 GCCCGGAGAAATGAGAAAATGGGGCCAGGGTGTCTTCTCCCTCGATTTGGAACCTG 8315
QY 143 ----- AspTyr 144
Db 8316 AACTGTCTTACCCCCATATCCCCGCTTTTTTTCCTTTTTTTTTTTTGAAGATTAT 8375
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164

Db 8376 TTTTACTGCTGAATACTTTTGTAGAAAACCCAGAAAGAACTTTTCAAAGCCTGGAGGG 8435
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181
Db 8436 CTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTG 8486
RESULT 5
US-09-966-880A-13
; Sequence 13, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-13
Alignment Scores:
Pred. No.: 9,39e-62 Length: 271
Score: 505.00 Matches: 90
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 46.50% Indels: 0
DB: 0 Gaps: 0
US-09-966-880A-8 (1-198) x US-09-966-880A-13 (1-271)
QY 53 AsnGlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72
Db 1 AACGGCTGCCAGCTGGAATGTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCT 60
QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92
Db 61 GGCGGCTGTACCGGCTCACCTGGTTCACCTCTCTGGAGCCCTGTCTACGACTGTGCCGA 120
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112
Db 121 CATGTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCC 180
QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAla 132
Db 181 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGGCGGCTGCACCGGCC 240
QY 133 GlyValGlnIleAlaIleMetThrPheLys 142
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAA 270
RESULT 6
US-09-729-674-173
; Sequence 173, Application US/09729674
; Patent No. US2001003935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.

```
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6053-64x
; CURRENT APPLICATION NUMBER: US/09/729,674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 173
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-729-674-173

Alignment Scores:
Pred. No.: 2,06e-44 Length: 1534
Score: 390.00 Matches: 83
Percent Similarity: 61.62% Conservative: 31
Best Local Similarity: 44.86% Mismatches: 59
Query Match: 35.91% Indels: 12
DB: 10 Gaps: 3

US-09-966-880A-8 (1-198) x US-09-729-674-173 (1-1534)

QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25
DB 639 ATGGATCCACCCCACTTCACTTCAACATTTAAACATGAACCTTGGGTCAGAGCGGCAT 698
QY 26 GluThrTyrLeuCysTyrValValLysArgAspSerAlaThrSerPheSerLeuAsp 45
DB 699 GAGACTTACTGTGTATGAGTGGAGCGCATGCACATGACACCTGGGTCCTGCTGAC 758
QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55
DB 759 CAGCGCAGGGCTTCTATGCAACAGCGCTCCACATAAACACGGTTTCTTGAAGGCCGC 818
QY 56 HisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75
DB 819 CATGCAGAGCTGTGCTCTCTGGACGTGATTCCCTTTTGGAAAGCTGGACCTGGACCAGGAC 878
QY 76 TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAla 95
DB 879 TACAGGGTTACTGTCTTCACTTCTGGAGCCCTGCTTCACTGTGCGCCAGGAATGGCT 938
QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115
DB 939 AAATTCATTTCATAAAACAAACACGCGGCTGTGCTATCTTCACTGCCCGCATCAT--- 995
QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135
DB 996 ---GATGATCAAGGAAGATGTCTCAGGAGGGGCTGCCACCTCTGCCCGAGCTGGGCCAAA 1052
QY 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155
DB 1053 ATTTCATATGACATACACTGATTAAGTACTGCTGGACACCTTTGTGGACCACCGAC 1112
QY 156 GluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
DB 1113 GGATGTCCTTCCAGCCCTGGGATGGATGAGTATGATGACACAGCAAGACCTGATGGGAGG 1172
QY 176 LeuArgArgIleLeu 180
DB 1173 CTGCGGGCCATTCTC 1187
```

RESULT 7

```
US-09-925-300-699
; Sequence 699, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 699
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-699
```

Alignment Scores:

```
Pred. No.: 7.95e-42 Length: 987
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 10 Gaps: 4
```

US-09-966-880A-8 (1-198) x US-09-925-300-699 (1-987)

```
QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArgGluThrTyrLeuCys 30
DB 147 TTCTACTTCCAAATTTAAAAACCTATATGGGAAGCCAGATCGGAAGCAAACTTTGGTGTGC 206
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
DB 207 TTCACCGTGAAGGTATTAAGCGCGCTCAGTTGTCTCTCTCGGAAGCG-----GGCGTC 260
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuPheLeuArgTyr 64
DB 261 TTCGGAACACAGGTGATTCAGACCCATTCTCATGAGAAAGGTGCTTCTCTCTCTTGG 320
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
DB 321 TTCTGCGACGACATCTGTCTCTTAACACAAAGTACCAGGTACCTGCTACACATCTTGG 380
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
DB 381 ACCCTTGGCCAGACTGTGACGGGAGGTGGCGGAGTTCTCTGGCCAGGCACAGCAGGTG 440
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
DB 441 AATCTCACCATCTTCCACCGCCGCTCTACTACTTCCAG---TATCCATGTTTACCAGGAG 497
QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
DB 498 GGGCTCCGCGAGCTGTAGTCAGGAAGGGTCTGCTGGAGATCATGACATATGAAGATTTT 557
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
DB 558 AAATATTTGTGGAAAACTTTGTGTACATATGATATGAGCCCATTCAGCCTTGGAGGGA 617
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 618 TTAATAAACCAACTTTTCGACTTCTGAAAGAGAGGCTACGGGAGAGTCTC 665
```

RESULT 8

```
US-10-098-841-268
; Sequence 268, Application US/10098841
```

Publication No. US20020197679A1

GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunqing
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhao, Qiang A.
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Qian, Xiaohong B.
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: NO. US20020197679A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2
CURRENT APPLICATION NUMBER: US/10/098,841
CURRENT FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598,042
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 331
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 268
LENGTH: 1143
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (128)..(700)
US-10-098-841-268

Alignment Scores:
Pred. No.: 9,98e-42 Length: 1143
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 9 Gaps: 4

US-09-966-880A-8 (1-198) x US-10-098-841-268 (1-1143)

Qy 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
Db 176 TTCACCTCCAAATTTAAACCACTATGGGAAGCCCAACGATCGGAACGAACTTGGCTGTC 235
Qy 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
Db 236 TTCACCGTGAAGGTATTAAGCGCGCTCAGTTGCTCTCTCGGAACAGC-----GGCGTC 289
Qy 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db 290 TTCGGAACCAAGGTGGATTCAGACACCATTTGTCATGCAAGAGGTGCTCTCTCTTGG 349
Qy 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
Db 350 TTCGCGAGCAGACTCTCTCTTAACCAAGTACCAGGTACCTGGTACACATCTTGG 409
Qy 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db 410 AGCCCTGCCAGACTGTGCAGGGGAGGTGCGGAGTTCCTGGCCAGCAGCAACGTG 469
Qy 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
Db 470 AATCTCACCATCTTCACCGCGCGCTCTACTACTTCCAG---TATCCATGTTACCAGGAG 526

Qy 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db 527 GGGCTCCGAGCCTGAGTCAGGAAGGGGTGCGTGTGGAGATCATGACTATGAAGATTTT 586
Qy 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db 587 AAATATTGTTGGGAAACCTTTCTGTACAAATGATAATAGAGCCATTCAAGCCTTGAAGGGA 646
Qy 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 647 TTAACCAACCACTTTTCGACTTCTGAAAGAGAGGCTACGGGAGAGTCTC 694

RESULT 9

US-09-880-107-3282
Sequence 3282, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3282
LENGTH: 1348
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891
NAME/KEY: unsure
LOCATION: (1)..(1348)
OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3282

Alignment Scores:
Pred. No.: 9e-41 Length: 1348
Score: 363.50 Matches: 83
Percent Similarity: 58.12% Conservative: 28
Best Local Similarity: 43.46% Mismatches: 57
Query Match: 33.47% Indels: 23
DB: 10 Gaps: 7

US-09-966-880A-8 (1-198) x US-09-880-107-3282 (1-1348)

Qy 5 LeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArg 24
Db 109 TTGATGGATCCACACATATTCACCTCCAACTTTAAACAAT-----GGCATTTGAAGG 159
Qy 25 ArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeu 44
Db 160 CATAAGACCTACTCTGTCTACGAAGTGGAGCGCTGGACAATGGCACCTCGGTCAAGATG 219
Qy 45 Asp-----PheGlyTyrLeuArgAsn-----LysAsn-----Gly 54
Db 220 GACCAAGCAGAGGGCTTTCTACACACACAGGCTAAGAATCTTCTGTGGCTTTTACGGC 279
Qy 55 CysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArg 74
Db 280 CGCCATGCGGAGCTCGGCTTCTTGGACCTGGTTCCTTTCAGTGTGACCGCGCCAG 339
Qy 75 CysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAsp-----CysAlaArg 92
Db 340 ATCTACAGGGTCACCTGGTTTCATCTCCTGGAGCCCTGCTTCTCTCTGGGCTGTGCGGG 399

```
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112
Db 400 GAAGTGGCGTTCCTTCAGGAGAACACACACGTCGAGCTGGCGATCTTCGTCGCCGC 459
QY 113 LeuTyrPheCysGluAspArgLysAlaGluPro-----GluGlyLeuArgArgLeu 129
Db 460 ATCTAT-----GATTACGACCCCTATATAAGGAGGCGCTGCATAATGCTG 504
QY 130 HisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsn 149
Db 505 CGGGATGCTGGGCGCCCAAGTCTCCATCATGACCTACGATGAATTTAAGCACTGCTGGGAC 564
QY 150 ThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSer 169
Db 565 ACCTTTGTGGACCAACACAGGATGTCCTCTCCAGCCCTGGGATGAGCTAGATGAGCACAGC 624
QY 170 ValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 625 CAAGCCCTGAGTGGGAGGCTGCGGGCCATTCTC 657

RESULT 10
; US-09-822-830A-359
; Sequence 359, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822.830A
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 359
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-359

Alignment Scores:
Pred. No.: 1.01e-32 Length: 2151
Score: 308.50 Matches: 81
Percent Similarity: 58.20% Conservative: 29
Best Local Similarity: 42.86% Mismatches: 60
Query Match: 28.41% Indels: 22
DB: 10 Gaps: 4

US-09-966-880A-8 (1-198) x US-09-822-830A-359 (1-2151)
QY 6 MetaAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25
Db 1278 ATGATCCACCCACATTCACTTCAACTTTAACAACTTGGTTCAGAGA-CGGCAT 1336
QY 26 GluThrTyrLeuCystYrValValLysArg-----AspPheGlyTyrLeuArgAsnLy 37
Db 1337 GAGACTTACCTGTGTATGAGTGGACGCGATGCAACATGACACCTGGTCCCTGCTGAA 1396
QY 38 SerAla-ThrSerPheSerLeu-----AspPheGlyTyrLeuArgAsnLy 52
Db 1397 CCAGCGCAGGGGCTTTCTTTATGACACCGAGGCTCCACATAAACACACGCTTTCCTT 1448
QY 52 sAsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeu 72
Db 1449 -GAAGGCCGCATGCAGAGCTGTCTTCCTGGAGCTGATCCCTTTTGGAGCTGGACCT 1507
```

```
QY 72 oGlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaAr 92
Db 1508 GGACCAGGACTACAGGTTTACCTGCTTACCTCTGGAGCCCTGCTTACGCTGTGCCCA 1567
QY 92 gHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaAr 112
Db 1568 GGAATGGCTAAATTCATTTCAAAAACAAACACACGTCGAGCTGTGCATCTTTCACCTGCCGC 1627
QY 112 qLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAl 132
Db 1628 CATCTAT-----GATGATCAAGGAAGATGTCAGAGGGGCTGCCACCTTGGCCGAGGC 1681
QY 132 aGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheVa 152
Db 1682 TGGGGCCAAAATTCATTAATGACATACAGTGAATTTAAGCACTGCTGGAC-ACCTTTGG 1740
QY 152 lGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLe 172
Db 1741 -GACCACAGGGATGTCCTTCCAGCCCTGGGATGGACTAGATGAGCACACGCAAGACCT 1799
QY 172 uSerArgGlnLeuArgArgIleLeu 180
Db 1800 GAGTGGGAGGCTGCGGGCCATTCTC 1824

RESULT 11
; US-09-966-880A-12
; Sequence 12, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966.880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 148
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-12

Alignment Scores:
Pred. No.: 1.07e-27 Length: 148
Score: 260.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.94% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-12 (1-148)
QY 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGly 23
Db 2 CTCCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAGGT 61
QY 24 ArgArgGluThrTyrLeuCystYrValValLysArgArgAspSerAlaThrSerPheSer 43
Db 62 CGGCGTGAGACCTACCTGCTGCTAGTACGTAGAGGCGGTGACAGTGTACATCTTTTCA 121
QY 44 LeuAspPheGlyTyrLeuArgAsnLys 52
Db 122 CTGGACTTGTGTTATCTTCGCAATAAG 148
```


Arg 25
|
GGAAT 386

Asp 45
||||
GGAC 443

Group 61

GATTC 503
 rTrp 80
 ||||
 CCTGG 563
 rGly 100
 CGAG 623
 rGlys 120
 AGAT 683
 nile 136
 : : :
 CATG 740
 sGlu 156
 AAGA 800
 nLeu 176

ngCTG 860

expressed s

FILE REFERENCE: 2760-103
CURRENT APPLICATION NUMBER: US/10/157.031
CURRENT FILING DATE: 2002-05-30
NUMBER OF SEQ ID NOS: 415
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 879
TYPE: DNA
ORGANISM: Homo sapiens
US-10-157-031-13

Alignment Scores:
Pred. No.: 1.15e-20 Length: 879
Score: 218.50 Matches: 44
Percent Similarity: 59.52% Conservative: 31
Best Local Similarity: 34.92% Mismatches: 42
Query Match: 20.12% Indels: 9
DB: 9 Gaps: 4

US-09-966-880A-8 (1-198) x US-10-157-031-13 (1-879)

```
QY 35 ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArg----- 50
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 119 CGTAAGAGCCCTGCTGCTCTACGAAATCAAGTGGGCAATGACCGGAAGATCTGCCGA 178

QY 51 -----AsnLysAsnGlyCys----HisValGluLeuLeuPheLeu---ArgTyrIleSer 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 AGCTCAGGCAAAAACACCACCAATCAGCTGGAAGTTAATTTATAAAAAAATTTACGTCA 238

QY 67 AspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrpSerPro 86
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 239 GAAAGAGATTTTCAACCATCCATCAGCTGCTCCATCACCCTGCTTGTCTCGAGTCCC 298

QY 87 CysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeu 106
|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 299 TGCTGGGAATGCTCCAGGCTATTAGAGAGTTCTGAGTCGGCACCTGGTGTGACTCTA 358

QY 107 ArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeu 126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 GTGATCTAGCTAGCTCGGCTTTTGGCACATGATGATCAACAA---AATCGCAAGGTCTC 415

QY 127 ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr 146
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 416 AGGACCTTGTAAACAGTGGTAATATTTCAGATTATGAGAGCATCAGAGTATTATCAC 475

QY 147 CysTrpAsnThrPheVal 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 476 TGCTGGAGGAATTTTGTCT 493
```

RESULT 15

US-09-966-880A-14
Sequence 14, Application US/09966880A
Patent No. US20020164743A1
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR APPLICATION NUMBER: JP 11-87192
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 116
TYPE: DNA

ORGANISM: Homo sapiens
US-09-966-880A-14

Alignment Scores:
Pred. No.: 4.12e-21 Length: 116
Score: 212.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.52% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-14 (1-116)

```
QY 144 TyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGlu 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 TATTTTACTGCTGGAATACTTTTGTAGAAAACCCAGAAAGAACTTTTCAAAAGCCCTGGAA 62

QY 164 GlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 63 GGGCTGCATGAAAATTTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116
```

Search completed: June 19, 2003, 19:06:49
Job time : 161 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:33:05 ; Search time 14079.9 Seconds
(without alignments)
11397.322 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagagcaatcatcgtgtcca.....tcaaaactcctgacctcagag 5514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	5482	99.4	71132	9	AC092184
2	4922.8	89.3	11204	9	AB040430 Homo sapi
3	251.8	4.6	155922	9	AL160291 Human DNA
4	243.2	4.4	161776	9	AL133229 Human DNA
5	242.2	4.4	184672	2	AF002508 Homo sapi
6	241.8	4.4	143517	9	AC083783 Homo sapi
7	241.8	4.4	198582	9	AC005291 Homo sapi
8	241.4	4.4	31632	9	AC104520 Homo sapi
9	241.4	4.4	41407	9	AC053467 Homo sapi
10	241.2	4.4	171073	2	AL590609 Homo sapi
11	241.2	4.4	270269	2	AL590728 Homo sapi
12	240.4	4.4	155085	2	AC025620 Homo sapi
13	240.4	4.4	186107	2	AC023132 Homo sapi
14	240.4	4.4	208164	2	AC113194 Homo sapi
15	239.4	4.3	155521	9	AC069335 Homo sapi
16	239.2	4.3	224048	2	AC023786 Homo sapi
17	239	4.3	84170	9	HS130N4
18	239	4.3	123501	9	AL390241 Human DNA
19	239	4.3	145166	2	AC053473 Homo sapi
20	239	4.3	164810	2	AC023411 Homo sapi
21	239	4.3	182653	9	AL354861 Human DNA
22	239	4.3	266079	2	AL391060 Homo sapi
23	238.8	4.3	66000	9	AL353665 Human DNA
24	238.6	4.3	172611	9	AC019129 Homo sapi
25	238.4	4.3	153665	2	AL138818 Homo sapi
26	236.6	4.3	192104	2	AC090229 Homo sapi
27	236.6	4.3	214978	9	AC007383 Homo sapi
28	236	4.3	167758	2	AC068746 Homo sapi
29	236	4.3	186299	2	AC107202 Homo sapi
30	236	4.3	219554	2	AC107016 Homo sapi
31	235.8	4.3	106896	9	AC074375 Homo sapi
32	235.4	4.3	133226	2	AL512364 Homo sapi
33	234.6	4.3	95719	2	AC015694 Homo sapi
34	234.6	4.3	180776	2	AC032043 Homo sapi
35	234.2	4.2	84570	9	AL359552 Human DNA
36	234.2	4.2	162964	9	AC011939 Homo sapi
37	234	4.2	158395	9	AC009242 Homo sapi
38	233.4	4.2	142281	9	HS4289880
39	233.4	4.2	176968	9	AC005669 Homo sapi
40	233.4	4.2	209772	9	AC099782 Homo sapi
41	233.2	4.2	85500	9	HS199H16 Human DNA
42	233	4.2	158838	9	AC009696 Homo sapi
43	232.6	4.2	176343	9	CNS01DX3
44	232.4	4.2	168297	2	AC011271 Homo sapi
45	232.4	4.2	170517	9	AC010545 Homo sapi

ALIGNMENTS

RESULT 1
AC092184
LOCUS AC092184 Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC092184 AC013443
VERSION AC092184.7 GI:21206067
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 71132)
AUTHORS Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
Alsbrooks S.L., Amaratunge H.C., Are J.R., Ayele M., Banks T.,

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouay, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carter, M., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrill, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooshtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kucherlapati, R., Weinstock, G. and Gibbs, R.

Direct Submission
2 (bases 1 to 71132)
Worley, K.C.
Submitted (25-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 71132)
Worley, K.C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 71132)
Worley, K.C.
Direct Submission
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 71132)
Worley, K.C.
Direct Submission
Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 25, 2002 this sequence version replaced gi:20901754.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	source	Location/Qualifiers
		1..71132
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="12"
		/clone="RP11-438L7"
		complement(1..1448)
		/notes="overlaps bases 1..1448 of clone AC092490"
		/function="clone overlap"
STS		187..408
		/standard_name="57233"
STS		439..560
		/standard_name="92005"
repeat_region		774..881
		/rpt_family="MIR"
repeat_region		903..1190
		/rpt_family="AluSq"
repeat_region		1191..1213
		/rpt_family="AT_rich"
STS		1591..1807
		/standard_name="6612"
STS		1744..1819
		/standard_name="8198"
repeat_region		1966..2264
		/rpt_family="AluSq"
repeat_region		3296..3328
		/rpt_family="(TTTC)n"
repeat_region		4097..4249
		/rpt_family="AluSq"
repeat_region		4360..4654
		/rpt_family="AluY"
repeat_region		complement(5140..5262)
		/rpt_family="FLAM_C"
repeat_region		5275..5304
		/rpt_family="AT_rich"
repeat_region		5305..5563
		/rpt_family="AluJo"
repeat_region		5564..5599
		/rpt_family="AT_rich"
repeat_region		5644..5733
		/rpt_family="MSTD"

Db 37619 AACGCTGTTTCTATAACAAAATTAGCCGGGATGGTGGCATGTGCCCTGTGTCCTCCAGCT 37678
 QY 3721 ACTAGGGGCTGAGCAGAGAGAACTTTGGAGCCAGGAGGTCAGGCTGCACTGAGCAG 3780
 Db 37679 ACTAGGGGCTGAGCAGAGAGAACTGGTGGAGCCAGGAGGTCAGGCTGCACTGAGCAG 37738
 QY 3781 TGCCTGGCCCACTGCACCTCCAGCCTGGGTGACAGGACGACCTTGCCTCAAAAAATAA 3840
 Db 37739 TGCCTGGCCCACTGCACCTCCAGCCTGGGTGACAGGACGACCTTGCCTCAAAAAATAA 37798
 QY 3841 GAAGAAAAATTAATAAATAAATGGAACCAACTACAAAGAGCTGTTGCTCTAGATGAGCTAC 3900
 Db 37799 GAAGAAAAATTAATAAATAAATGGAACCAACTACAAAGAGCTGTTGCTCTAGATGAGCTAC 37858
 QY 3901 TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGCACCTGCACATAC 3960
 Db 37859 TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGCACCTGCACATAC 37918
 QY 3961 ATTATTAATAATCAATTTCTCAATGTATATCCACACAAAGACTGGTACGTGATGTTTCAAT 4020
 Db 37919 ATTATTAATAATCAATTTCTCAATGTATATCCACACAAAGACTGGTACGTGATGTTTCAAT 37978
 QY 4021 AGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA 4080
 Db 37979 AGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA 38038
 QY 4081 CAAATAACAAAATGTGCTATATCCATGCAATGGAATACCAACCTGCAGTACAAAGGAAG 4140
 Db 38039 CAAATAACAAAATGTGCTATATCCATGCAATGGAATACCAACCTGCAGTACAAAGGAAG 38094
 QY 4141 AAGCTACTTGGGATGAATCCCAAGTATGACGCTTAATGAAAGACTCAGACATGAAGG 4200
 Db 38095 AAGCTACTTGGGATGAATCCCAAGTATGACGCTTAATGAAAGACTCAGACATGAAGG 38154
 QY 4201 AGGAGATAATGTATGCCATACGAAATCTAGAAAATGAAAGTAACCTATATGTTACAGAAA 4260
 Db 38155 AGGAGATAATGTATGCCATACGAAATCTAGAAAATGAAAGTAACCTATATGTTACAGAAA 38214
 QY 4261 GCAATACGGCAGGCATPAGAGGCTCAGCCTGTAATCCAGCACTTTGAGAGGCCACGT 4320
 Db 38215 GCAATACGGCAGGCATPAGAGGCTCAGCCTGTAATCCAGCACTTTGAGAGGCCACGT 38274
 QY 4321 GGGAGATTGCTAGAACTCAGAGGTTCAAGACCAGCCTGGGCAACACAGTGAACCTCCAT 4380
 Db 38275 GGGAGATTGCTAGAACTCAGAGGTTCAAGACCAGCCTGGGCAACACAGTGAACCTCCAT 38334
 QY 4381 TCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTCTGTGGGAGGGGAAG 4440
 Db 38335 TCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTCTGTGGGAGGGGAAG 38394
 QY 4441 GACTGCCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGGTGAATTCAGGTTCTGTATCCT 4500
 Db 38395 GACTGCCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGGTGAATTCAGGTTCTGTATCCT 38454
 QY 4501 GACTGTGCTAGCAGTTTGGGCTGTTTACATCCAAAATATTCGTAGAAATTTATGCATCTTA 4560
 Db 38455 GACTGTGCTAGCAGTTTGGGCTGTTTACATCCAAAATATTCGTAGAAATTTATGCATCTTA 38514
 QY 4561 AATGGTGGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAAATAATGTTAAG 4620
 Db 38515 AATGGTGGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAAATAATGTTAAG 38574
 QY 4621 AAAAGTTCAATTTCTTTGCCAGCAACGTTTATCAAAATCCTGAGCCCTTTACTTCGCA 4680
 Db 38575 AAAATTTTCAATTTCTTTGCCAGCAACGTTTATCAAAATCCTGAGCCCTTTACTTCGCA 38634
 QY 4681 AATCTCTGCACCTTCTGCCCGTACCATTAGGTGACACACACTAGCTCCACAAATGGATA 4740
 Db 38635 AATCTCTGCACCTTCTGCCCGTACCATTAGGTGACACACACTAGCTCCACAAATGGATA 38694
 QY 4741 AATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATCAGTTGTGCTTTTCATATCAA 4800
 Db 38695 AATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATCAGTTGTGCTTTTCATATCAA 38754

QY 4801 CCACGCTGTACAGCTTGTGTGCTGTCTCAGCTGCAATGGGGACCTTGTGATTTCTTTAA 4860
 Db 38755 CCATGCTGTACAGCTTGTGTGCTGTCTCAGCTGCAATGGGGACCTTGTGATTTCTTTAA 38814
 QY 4861 GCAGAACTTGGGTTACAGAGTATTTCCACAAATGCTATTCAAAATAGTCTGTATGATG 4920
 Db 38815 GCAGAACTTGGGTTACAGAGTATTTCCACAAATGCTATTCAAAATAGTCTGTATGATG 38874
 QY 4921 CAAGACACTGTCTAGGAGCCAGAAAAACAAGAGGAGAGAAATCAGTATTATGTGGGA 4980
 Db 38875 CAAGACACTGTCTAGGAGCCAGAAAAACAAGAGGAGAGAAATCAGTATTATGTGGGA 38934
 QY 4981 ACAACATAGCAAGATATTTAGATCATTTTACTAGTATTAAAAAGCAGCAGAGTACAAAAT 5040
 Db 38935 ACAACATAGCAAGATATTTAGATCATTTTACTAGTATTAAAAAGCAGCAGAGTACAAAAT 38994
 QY 5041 CACACATGCAATCAGTATTAATCCAAATCATGTAAATATGTGCTGTAGAAAGACTAGAGG 5100
 Db 38995 CACACATGCAATCAGTATTAATCCAAATCATGTAAATATGTGCTGTAGAAAGACTAGAGG 39054
 QY 5101 AATAAACACAAGAATCTTAAACAGTCAATGCTCAATAGACACTAAGTCTAAATATTATTATT 5160
 Db 39055 AATAAACACAAGAATCTTAAACAGTCAATGCTCAATAGACACTAAGTCTAAATATTATTATT 39114
 QY 5161 AGACACTATGATATTTGAGATTTAAAAAATCTTTAAATATTTTAAAAATTTAGAGCTCTTCT 5220
 Db 39115 AGACACTATGATATTTGAGATTTAAAAAATCTTTAAATATTTTAAAAATTTAGAGCTCTTCT 39174
 QY 5221 ATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTT 5280
 Db 39175 ATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTT 39234
 QY 5281 TTTTCTTTTTTTTGGAGATGGAGTTTGGTCTTTGTTGCCCATGCTGGAGTGGAAATGGCATG 5340
 Db 39235 TTTTCTTTTTTTTGGAGATGGAGTTTGGTCTTTGTTGCCCATGCTGGAGTGGAAATGGCATG 39294
 QY 5341 AYCATAGTCTACTGCAACCTCCACCTCTCGGTTCAAGCAAAAGCTGTGCGCTCAGCCTCC 5400
 Db 39295 ACCATAGTCTACTGCAACCTCCACCTCTCGGTTCAAGCAAAAGCTGTGCGCTCAGCCTCC 39354
 QY 5401 CGGTAGATGGATACAGGCCGCCACCAACACACACTCGGCTAATGTTGTATTTTAGTA 5460
 Db 39355 CGGTAGATGGATACAGGCCGCCACCAACACACACTCGGCTAATGTTGTATTTTAGTA 39414
 QY 5461 GAGATGGGTTTCCACATGTTGGCAGGCTGTCTCAAACTCCTGACCTCAGAG 5514
 Db 39415 GAGATGGGTTTCCACATGTTGGCAGGCTGTCTCAAACTCCTGACCTCAGAG 39468

RESULT 2
 AB040430
 LOCUS Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.
 DEFINITION
 ACCESSION AB040430
 VERSION AB040430.1 GI:9988407
 KEYWORDS AID; activation-induced cytidine deaminase.
 SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.
 TITLE Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene
 JOURNAL Genomics 66 (1), 85-88 (2000)
 MEDLINE 20408890
 REFERENCE 2 (sites)
 AUTHORS Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., Tescan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A.

TITLE	and Durandy,A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-Igm syndrome (HIGM2)					
JOURNAL	Cell 102 (5), 565-575 (2000)					
REFERENCE	20460541					
AUTHORS	3 (bases 1 to 11204) Muto,T., Muramatsu,M., Tanikawa,M., Kinoshita,K. and Honjo,T.					
JOURNAL	Direct Submission Submitted (18-Mar-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan [E-mail:honjo@four.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388]					
FEATURES	Location/Qualifiers 1..11204					
source	/organism="Homo sapiens" /db_xref="taxon:9606"					
gene	join(521..528,6280..6427,7807..8077,8371..8486,8956..9009) /gene="AID".528,6280..6427,7807..8077,8371..8486,8956..9009; /gene="AID"					
CDS	join(521..528,6280..6427,7807..8077,8371..8486,8956..9009) /codon_start=1 /product="activation-induced cytidine deaminase" /protein_id="BAB12720.1" /db_xref="GI:9988408"					
	/translation="MDSLUMNRKFLYQFKVNRWAKGRRETYLYCVVKRRDSSATSFSL DFGLRNKGCHVELLFYITSDMDLPGRCYRVTFWSWPCYDCARHVADFLRGNP NLSRIETARLYFCEDRAKEPELRRLHRAGVOIAIMTFKDIFYCWNFTVENHERTEFK AWEGLEHSNVRLSRQLRILLPLYEVDDLDAFRTLLG"					
BASE COUNT	3305 a 2273 c 2373 g 3253 t					
ORIGIN						
Query Match	89.3%; Score 4922.8; DB 9; Length 11204;					
Best Local Similarity	99.9%; Pred. No. 0;					
Matches 4921; Conservative	3; Mismatches 0; Indels 0; Gaps 0;					
Qy	591	AGGTTCAGAGACTGTGGGAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACAC	650			
Dd	1	AGGTTCAGAGAGACTGTGGGAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACAC	60			
Qy	651	AATAACCCAAAGAAGCTATTTAAATGCCTCTTAAAGTGATTTACATAAATATTACTATTCTC	710			
Dd	61	AATAACCCAAAGAAGCTATTTAAATGCCTCTTAAAGTGATTTACATAAATATTACTATTCTC	120			
Qy	711	ATTGTGCTTTTATTTTGTGTATCATGATTAATGAAGTGCTCTACTGTACTGCCCTCC	770			
Dd	121	ATTGTGCTTTTATTTTGTGTATCATGATTAATGAAGTGCTCTACTGTACTGCCCTCC	180			
Qy	771	TGATCTTTGCTAGCTATGGAGCATGAGCTGGCTTTTTAGAGCAGCAGCCCCAAAGAAC	830			
Dd	181	TGATCTTTGCTAGCTATGGAGCATGAGCTGGCTTTTTAGAGCAGCAGCCCCAAAGAAC	240			
Qy	831	TAAACATTAAAGCAGAGCTGCCCAATGGTTTAACTCTGTGACTCTGCCTATCACAGC	890			
Dd	241	TAAACATTAAAGCAGAGCTGCCCAATGGTTTAACTCTGTGACTCTGCCTATCACAGC	300			
Qy	891	CCCACCACCCATCTTCACTGGATCCAAATCAGGAGCAAGCGCTGGGGTACCTGGTGG	950			
Dd	301	CCCACCACCCATCTTCACTGGATCCAAATCAGGAGCAAGCGCTGGGGTACCTGGTGG	360			
Qy	951	GGGTGATGCTCTCAGGGAGGAGCCCCAAAAGGGCAAGCTCAAATTTGAATGTGAAGGCC	1010			
Dd	361	GGGTGATGCTCTCAGGGAGGAGCCCCAAAAGGGCAAGCTCAAATTTGAATGTGAAGGCC	420			
Qy	1011	AATSCACTGTTCAGACTGNAGACAGAACCATCATTAATTGAATGAGTAGATTTTCTGGCGCT	1070			
Dd	421	AATSCACTGTTCAGACTGNAGACAGAACCATCATTAATTGAATGAGTAGATTTTCTGGCGCT	480			
Qy	1071	GAGACTTGCAGGGAGGCAAGAACACACTCTCGACACCACTATGGACAGGTAAAGAGGCAG	1130			
Dd	481	GAGACTTGCAGGGAGGCAAGAACACACTCTCGACACCACTATGGACAGGTAAAGAGGCAG	540			
Qy	1131	TCCTTCTCGTGGGTGATTGCAGCTGGCCCTTCCTCTCAGAGCAAACTCTGAGTAATGAGACTGG	1190			

QY 2271 TCAAAGTAAATGGAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAGACAGAAA 2330
DB 1681 TCAAGTAAATGGAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAGACAGAAA 1740
QY 2331 AGGAGAAAGATGAAATTCACAGACAGAGAGGAGGAAATATATATCATTAAGGAGGACAG 2390
DB 1741 AGGAGAAAGATGAAATTCACAGACAGAGAGGAGGAAATATATATCATTAAGGAGGACAG 1800
QY 2391 TATCTGTAGAGCTCATTAGTGTATGGCAAAATGACTTGTTCAGGATATATTTTAACCCGCT 2450
DB 1801 TATCTGTAGAGCTCATTAGTGTATGGCAAAATGACTTGTTCAGGATATATTTTAACCCGCT 1860
QY 2451 TGTCTCTGGTTTGCACGGCTGGGATGCAGCTAGGGTTCCTGCCCTCAGGAGACAGAGCTGT 2510
DB 1861 TGTCTCTGGTTTGCACGGCTGGGATGCAGCTAGGGTTCCTGCCCTCAGGAGACAGAGCTGT 1920
QY 2511 CCAGAGAGCTGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCCTACTPCA 2570
DB 1921 CCAGAGAGCTGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCCTACTPCA 1980
QY 2571 GGACAGAAATCAGCAGACAGGAGCTGGAACAGGCCCTTAACAGAGAAGGAGTAA 2630
DB 1981 GGACAGAAATCAGCAGACAGGAGCTGGAACAGGCCCTTAACAGAGAAGGAGTAA 2040
QY 2631 TGGATCAACAAAGTTAACTAGCAGTCAAGGATCAGCAATTCATTTCACTCTGACTGGTA 2690
DB 2041 TGGATCAACAAAGTTAACTAGCAGTCAAGGATCAGCAATTCATTTCACTCTGACTGGTA 2100
QY 2691 ACATGTGACAGAAACAGTGTAGGCTTATGTTATTTTCATGTAGAGTAGGACCCCAAAATC 2750
DB 2101 ACATGTGACAGAAACAGTGTAGGCTTATGTTATTTTCATGTAGAGTAGGACCCCAAAATC 2160
QY 2751 CACCCAAAGTCCCTTATCTATGCCCACATCCTTCTTATCTATCTTCCAGGACACTTTTC 2810
DB 2161 CACCCAAAGTCCCTTATCTATGCCCACATCCTTCTTATCTATCTTCCAGGACACTTTTC 2220
QY 2811 TTCCTTATGATGAAGCTCTCTCTCTCCACACACACACACACACACACACACACACACAC 2870
DB 2221 TTCCTTATGATGAAGCTCTCTCTCTCCACACACACACACACACACACACACACACACAC 2280
QY 2871 CAC 2930
DB 2281 CAC 2340
QY 2931 CTCTGCCTTTCTCATCTACACAGCCCGGAGGGTAAAGTTATATATAAGAGGATTTATTGG 2990
DB 2341 CTCTGCCTTTCTCATCTACACAGCCCGGAGGGTAAAGTTATATATAAGAGGATTTATTGG 2400
QY 2991 TAAGAGATGATGCTTAACTCTTTAACTGGGCTCAAGAGAGAAATTTCTTTCTTTCT 3050
DB 2401 TAAGAGATGATGCTTAACTCTTTAACTGGGCTCAAGAGAGAAATTTCTTTCTTTCT 2460
QY 3051 GTACTTATTAAGCACCTATTATGTGTGAGCTTATATATACAAAGGGTTATTATATGCTA 3110
DB 2461 GTACTTATTAAGCACCTATTATGTGTGAGCTTATATATACAAAGGGTTATTATATGCTA 2520
QY 3111 ATATAGTAAATAGTAAATGKTGGTGGTACTATGTTAAATACCATAAAATTAATATCCCTTT 3170
DB 2521 ATATAGTAAATAGTAAATGKTGGTGGTACTATGTTAAATACCATAAAATTAATATCCCTTT 2580
QY 3171 TAAATTAAGCTAAATATATTATGATCTTTTATGATTTATGTTATGTTTATGTTTATGTTT 3230
DB 2581 TAAATTAAGCTAAATATATTATGATCTTTTATGATTTATGTTTATGTTTATGTTTATGTTT 2640
QY 3231 TTGATTTTTTAAAGACAAATCTACCCCTGTTTACCCAGGCTGGAGTGCAGTGGTGAATPCA 3290
DB 2641 TTGATTTTTTAAAGACAAATCTACCCCTGTTTACCCAGGCTGGAGTGCAGTGGTGAATPCA 2700
QY 3291 TAGCTTTCTGAGGCTTTGAATCTCTGGGCTCAAGCAATCTCTCTGCTTGGCTTGGCTCCCAAA 3350
DB 2701 TAGCTTTCTGAGGCTTTGAATCTCTGGGCTCAAGCAATCTCTCTGCTTGGCTTGGCTCCCAAA 2760

QY 3351 GTGTTGGATACAGTACAGTGCATGAGCCACTGCATCTGGCTAGATCCATTTAGATTAATAATAT 3410
DB 2761 GTGTTGGATACAGTACAGTGCATGAGCCACTGCATCTGGCTAGATCCATTTAGATTAATAATAT 2820
QY 3411 GCATTTTAAATTTTAAATAATATATGGCTAAATTTTACCTTATGTATGTGTATACPGTGA 3470
DB 2821 GCATTTTAAATTTTAAATAATATATGGCTAAATTTTACCTTATGTATGTGTATACPGTGA 2880
QY 3471 ATAAATCTAGTGTGCTGCTTAAAGTAAAGTGCCTTCCAAATAGCTTTCATGCTGAG 3530
DB 2881 ATAAATCTAGTGTGCTGCTTAAAGTAAAGTGCCTTCCAAATAGCTTTCATGCTGAG 2940
QY 3531 GGGAGACATTTAAAGTGAACAGACAGCAGGCTGTGGCTCAGCCCTGTAAATCCAGC 3590
DB 2941 GGGAGACATTTAAAGTGAACAGACAGCAGGCTGTGGCTCAGCCCTGTAAATCCAGC 3000
QY 3591 ACTCTGGAGGCTGAGGTGGGTGGATCTTGAAGCCCTGGAGTTCAAAGACCAGCCTGAGC 3650
DB 3001 ACTCTGGAGGCTGAGGTGGGTGGATCTTGAAGCCCTGGAGTTCAAAGACCAGCCTGAGC 3060
QY 3651 AACATGGCAAAACCCCTGTTCTATTAACAAAATTTAGCCGGCATGGTGGCATGTGCTGT 3710
DB 3061 AACATGGCAAAACCCCTGTTCTATTAACAAAATTTAGCCGGCATGGTGGCATGTGCTGT 3120
QY 3711 GGTCCAGCTACTAGGGGCTGAGGACAGGAGAAATCTTTGGAGCCAGAGGCTCAAGGCTG 3770
DB 3121 GGTCCAGCTACTAGGGGCTGAGGACAGGAGAAATCTTTGGAGCCAGAGGCTCAAGGCTG 3180
QY 3771 CACTGAGCAGTCTTGGCCCACTGCACCTCCAGCCCTGGGTGACAGACCAGCCTTGGCTC 3830
DB 3181 CACTGAGCAGTCTTGGCCCACTGCACCTCCAGCCCTGGGTGACAGACCAGCCTTGGCTC 3240
QY 3831 AAAAAATTAAGAGAGAAATTTAAATAATTAATGGAACAACTACAAGAGCTGTGCTCTA 3890
DB 3241 AAAAAATTAAGAGAGAAATTTAAATAATTAATGGAACAACTACAAGAGCTGTGCTCTA 3300
QY 3891 GATGAGCTACTTAGTGTAGCTGATATTTGGTATTTAACTTTTAAAGTCAGGCTGTGCTCA 3950
DB 3301 GATGAGCTACTTAGTGTAGCTGATATTTGGTATTTAACTTTTAAAGTCAGGCTGTGCTCA 3360
QY 3951 CCTGCATCTACATTTTAAATAATATCAATTTCTCAATGTATATCCACACAAAGCTGTGCT 4010
DB 3361 CCTGCATCTACATTTTAAATAATATCAATTTCTCAATGTATATCCACACAAAGCTGTGCT 3420
QY 4011 GAATGTTCTATAGTACTTTATTCACAAACCCCAAGTAGACATATCCAAATATCCATC 4070
DB 3421 GAATGTTCTATAGTACTTTATTCACAAACCCCAAGTAGACATATCCAAATATCCATC 3480
QY 4071 AACAGTGAACAAATTAACAAATGTCTATATCCATGCAATGGAATACACCCCTGCTGCT 4130
DB 3481 AACAGTGAACAAATTAACAAATGTCTATATCCATGCAATGGAATACACCCCTGCTGCT 3540
QY 4131 ACAAGGAGAGCTTACTTGGGATGAATCCCAAGCTCATGACGCTTAAATGAAGAGTCA 4190
DB 3541 ACAAGGAGAGCTTACTTGGGATGAATCCCAAGCTCATGACGCTTAAATGAAGAGTCA 3600
QY 4191 GACATGAAGGAGGAGATTAATGTATGCCATACGAAATTTCTAGAAAATGAAGTAACTTATA 4250
DB 3601 GACATGAAGGAGGAGATTAATGTATGCCATACGAAATTTCTAGAAAATGAAGTAACTTATA 3660
QY 4251 GTTACAGAAAGCAATCAGGCGCAGGCTAGAGGCTCAGACCTGTAAATCCAGCAGCTTCA 4310
DB 3661 GTTACAGAAAGCAATCAGGCGCAGGCTAGAGGCTCAGACCTGTAAATCCAGCAGCTTCA 3720
QY 4311 GAGGCCAGTGGGAGATTGCTAGAACTCAGGAGTTCAGGACCCGCTGGGCAACACAGT 4370
DB 3721 GAGGCCAGTGGGAGATTGCTAGAACTCAGGAGTTCAGGACCCGCTGGGCAACACAGT 3780
QY 4371 GAAACTCCATTTCTCCACAAAATGGGAAAAAAGAAAGCAATAGTGGTGTCTGTGG 4430
DB 3781 GAAACTCCATTTCTCCACAAAATGGGAAAAAAGAAAGCAATAGTGGTGTCTGTGG 3840
QY 4431 GGAGGGAGGAGCTCAAGAGGGAGAGGCTCTGTTGGGTGAGGGTGTGCTTATTCAGCT 4490

Db 3841 GGAGGGGAAGGACTGCAAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGTGATTGAGGT 3900
QY 4491 TCTGTATCTGACTGTGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGCAATT 4550
Db 3901 TCTGTATCTGACTGTGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGCAATT 3960
QY 4551 ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATTTATACCTCAATGTAAGAAAAAT 4610
Db 3961 ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATTTATACCTCAATGTAAGAAAAAT 4020
QY 4611 AATGTGTAGAAAAAGTTTCAATCTCTGCCAGCAAAACGTTTATTCAAATTCCTGAGCCCT 4670
Db 4021 AATGTGTAGAAAAAGTTTCAATCTCTGCCAGCAAAACGTTTATTCAAATTCCTGAGCCCT 4080
QY 4671 TTACTTGGCAAAATCTCTGCACCTTCTGCCCTTACCATTAGTGTGACGACACTAGCTCCAC 4730
Db 4081 TTACTTGGCAAAATCTCTGCACCTTCTGCCCTTACCATTAGTGTGACGACACTAGCTCCAC 4140
QY 4731 AAATGGATTAATGCAATTTCTGAAAGACTAGGGACAAAATCCAGGCATCACTTTGTGCT 4790
Db 4141 AAATGGATTAATGCAATTTCTGAAAGACTAGGGACAAAATCCAGGCATCACTTTGTGCT 4200
QY 4791 TTCAATCAACACGCTGTAGAGCTTGTGCTGTCTGCAGCTGCAATGGGAGCTCTTG 4850
Db 4201 TTCAATCAACACGCTGTAGAGCTTGTGCTGTCTGCAGCTGCAATGGGAGCTCTTG 4260
QY 4851 ATTTCTTTAAGAAACTTGGGTTACAGAGTATTTCCACAAATGCTATTCAAAATAGTGC 4910
Db 4261 ATTTCTTTAAGAAACTTGGGTTACAGAGTATTTCCACAAATGCTATTCAAAATAGTGC 4320
QY 4911 TTATGATATGCAAGACACTGTCTAGGAGCCAGAAAACAAAGAGGAGGAGAATCAGTCA 4970
Db 4321 TTATGATATGCAAGACACTGTCTAGGAGCCAGAAAACAAAGAGGAGGAGAATCAGTCA 4380
QY 4971 TTATGAGGAACAATAGCAGATATTTAGATCATTTTGTGACTAGTTTAAAAAGCAGCAG 5030
Db 4381 TTATGAGGAACAATAGCAGATATTTAGATCATTTTGTGACTAGTTTAAAAAGCAGCAG 4440
QY 5031 AGTACAAAATCACATGCAATGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 5090
Db 4441 AGTACAAAATCACATGCAATGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 4500
QY 5091 AGACTAGAGGAATAAACACAGAATCTTAACAGTCAATGTGCTAATAGACACTAAGTCTAAT 5150
Db 4501 AGACTAGAGGAATAAACACAGAATCTTAACAGTCAATGTGCTAATAGACACTAAGTCTAAT 4560
QY 5151 TATTATTATTAGACACTATGATATTTAGATTTTAAATCTTTAATTTTAAATTTA 5210
Db 4561 TATTATTATTAGACACTATGATATTTAGATTTTAAATCTTTAATTTTAAATTTA 4620
QY 5211 GAGCTCTCTATTTTCCATAGTATTCAGTTTGAATGATCAAGTATTAATCTCTCTCTT 5270
Db 4621 GAGCTCTCTATTTTCCATAGTATTCAGTTTGAATGATCAAGTATTAATCTCTCTT 4680
QY 5271 TT 5330
Db 4681 TT 4740
QY 5331 GAATGGCATGAYCATGCTACTGCAACCTCCACCTCTGGTTCACGAAGCTGTGCG 5390
Db 4741 GAATGGCATGAYCATGCTACTGCAACCTCCACCTCTGGTTCACGAAGCTGTGCG 4800
QY 5391 CTCAGCCCTCCGGGTAGATGGGATTACAGGGGCCCCACCACACACTCGGCTAATGTTGT 5450
Db 4801 CTCAGCCCTCCGGGTAGATGGGATTACAGGGGCCCCACCACACACTCGGCTAATGTTGT 4860
QY 5451 ATTTTATTAGATAGGAGTTTACCATGTTGGCCAGCGTGTCTCAAACTCCTGACCTC 5510
Db 4861 ATTTTATTAGATAGGAGTTTACCATGTTGGCCAGCGTGTCTCAAACTCCTGACCTC 4920
QY 5511 AGAG 5514
|||||

Db 4921 AGAG 4924

RESULT 3
LOCUS AL160291
DEFINITION Human DNA sequence from clone RP11-85G18 on chromosome 10, complete sequence.
ACCESSION AL160291
VERSION AL160291.30
KEYWORDS GI:212111640
SOURCE HTG.
ORGANISM human.
REFERENCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155922)
Dunn, M.
Direct Submission
Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:18958078.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-85G18 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
Location/Qualifiers
1..155922
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-85G18"
/clone_lib="RPCI-11.1"

BASE COUNT 45636 a 32268 c 31587 g 46431 t
ORIGIN

Query Match 4.6%; Score 251.8; DB 9; Length 155922;
Best Local Similarity 50.3%; Pred. No. 6.4e-99;
Matches 1030; Conservative 1; Mismatches 943; Indels 73; Gaps 14;

QY 3539 TTAAAGTGAACAGACAGACAGCCAGGTGTGGTGCTCACGCTGTAAATCCAGCAGCTCTGGG 3598
|||||
Db 108116 TTTAAATTTTAAATTTGCCAGTTGTGGTGCTCATGTCTAATCCAGCAGCTTTGTG 108175
QY 3599 AGGCTGAGTGGGTGGATCGCTTTGAGCCCTGAGGTTCAAGACAGCCTGAGCAACATGCG 3658
|||||
Db 108176 AGGCTGAGTGGGTGGATCACCTGACCCAGGAGTTCAAGGCGAGCCTGGGCAACATAGC 108235
QY 3659 AAAACCCCTGTTTCT-----ATAACAAAAATAGCCGGCATGTGGCATGTGCTGTG 3711
|||||
Db 108236 AAAACCCCTGTTTCTACAAATAAAAAAATAGTGTGGTGTGATGGCA-CACCCCTGTA 108294

[illegible]

Qy	4731	AAATTGGATTAATGCAATTCCTGGAAAAGACACTAGGACAAAATCCAGGCATCACTTGTGCT	4790
Db	109374	CAGAGGGAAAAAATGGAGAAACACACACTTTTAAAATGGTCTACTTATTTTAAATAACACAAA	109433
Qy	4791	TTCATATCAACACGCTGTACAGCTTGTGGTCTCTGCTGACGTCGCAATGGGACTCTTG	4850
Db	109434	TCTAAATTTTAAACATCAAAAGTCTAATCTCAAAATATCTTCAGTAAGTCAACCAATAT	109493
Qy	4851	ATTTCCTTTAAAGAAACTTGGGTTTACCAGAGTATTTCCACAAATGCTATTTCAAAATPAGTC	4910
Db	109494	GTTTATTGGCTTAGGTGATGATATTAATAACAAAACCGAGAACATTTATTTCAGTGT	109553
Qy	4911	TTATGATATGCAAGACACTGTGCTAGGACCCAGAAAACAAAGAGGAGGAGAGAATCA--GT	4968
Db	109554	AGTCTGTGTGAAGACACTCAAGTAAGATTTCTGCTCAAGTGCTCACACATGATGTA	109613
Qy	4969	CATTATGTGGGAACAACATACGAAGATATTTAGATCATTTTTCAGCTAGTTTAAAAAGCAGC	5028
Db	109614	CTTCATATGATTTTGAATATTTAGCTAAGAAATTTTAAATTTTAAAGGTTTAAAGTCATACC	109673
Qy	5029	AGAGTACAAAATCACATGCAATCAGTATTAATCCAAATCATGTAAATAT--GTGCTCTGTA	5087
Db	109674	AAACATAGATGTCATAGGATTCATTACTATTTTCAGGCCCTTAATAAGGATTGAGAAACAA	109733
Qy	5088	GAAGACTAGAGGAATTAACACAGAATCTTTAACAGTCATTTGTCTATAGACACTAAGTCT	5147
Db	109734	GGACAAAAAACTCCAAGATCTTGCTTTGCTAATGATTCAGAGCAAAATCAAAATGGTT	109793
Qy	5148	AATTATTTATTTAGACACTATGATATTTGAGATTTAAAAATCTTTTAATATTTTAAAT	5207
Db	109794	TAAAGTAGTATTAACAAAAATTTTATATATCCATTAATCAAACTCTGCTTCAATTAAT	109853
Qy	5208	TTAGAGCTCTCTATTTTTCATAGTATTTCAAGTTTGACAATGATCAAGTATTACTCTTT	5267
Db	109854	ATCAACTTAGTTTCAGGGATACAGTATATGATGTTTCCACACACAGTACCATTTG	109913
Qy	5268	CTTTTTTTTTTTTTTTTTTTTTTTTGGAGTGGAGTTTTGGTCTTTGGCCCATGCTGGA	5327
Db	109914	ATAGGATATCTTTTTTTTTTTTTTTTTTTTGGAGTGGAGTCTGGCTTGTCAACCCAGCTGA	109973
Qy	5328	GTGAATGGCATGAYCATGCTACTGCAACTCCACCTCTGGGTTCAAGCAAAAGCTGT	5387
Db	109974	GTCAATATGGCGCAATCTCAGCTGACTGCAACTCCGCTCCGAGTTCAAGGATTTTCC	110033
Qy	5388	CGCTCAGCCCTCCGGGTAGATGGGATTTACAGCGCCACCACCACTCGGCTAATGTT	5447
Db	110034	TGCTCAACCTCCCGAGTAGTGGGATTTACAGGAACCTGCCACTACACTACGCTAATTT	110093
Qy	5448	TGTATTTTATAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC	5507
Db	110094	TGTATTTTATAGAAAGTGGGTTTTCACCATGTTGGCCAGGCTGGTCTGGAATTCCTGAT	110153
Qy	5508	CTCAGAG 5514	
Db	110154	CTCAGGG 110160	
RESULT 4			
AL133229			
LOCUS	161776 bp DNA linear PRI 06-APR-2001		
DEFINITION	Human DNA sequence from clone Rps-1167E19 on chromosome 20 Contains STSs, GSSs and a CpG island, complete sequence.		
ACCESSION	AL133229		
VERSION	AL133229.40 GI:8546579		
KEYWORDS	HTG; CpG island.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 161776)		
JOURNAL	Direct Submission		
	Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,		

[illegible]

JOURNAL

Db 30596 AAGTGAAT-----TGGCAAGAGCACTTTGGAGAGCTCAGCAAGCAATATCTAAAG 30646
Qy 4848 TTGATTTCTTTAAGAAACTTGGGTTACCAGAGATTTTCCACAAATCTTATCAAAATAG 4907
Db 30647 CTCAGGTGCTTATGCTCCTCACTCAGACTCAGCACTTCTACATG-----TCAAGAGG 30702
Qy 4908 TGCTTATGATGATGAGAGCACTGTGTGAGGAGCAAAACAAAGAGGAGGAGAAATCAG 4967
Db 30703 AGACATTTGCAAGAAAGTTCATTGTAACATTTGTTGTAACAATAAAAAAGAGAAAAATC 30762
Qy 4968 TCATTATGTGGGAACACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAGCAG 5027
Db 30763 ACCCAATGTCACCAACAGGCAATGGGTACATACATTTGTGTATATTCATATATAG 30822
Qy 5028 CAGAGTACAAATCACACATGCAATCAGTATATTCATAATCCAAATCTGTAATATGTCCTGTA 5087
Db 30823 AAAAAAGAAATTAATGAACCAACAAAGAAATTAATGTACACAAAGAAATTAATGAA 30882
Qy 5088 GAAAGACTAGAGAAATAACA-CAAGATCTTAAACAGTCAATGCTCATTAGACACTAAGTC 5146
Db 30883 CCACAGCTAGATGTCTATGATTAATGTTACTGTCAAAAACATAATGTTAAAGTTAAAAAAC 30942
Qy 5147 TAATTATTTATTAGACACTATGATATTTGAGATTTAAAAATCTTTAATATTTAAAA 5206
Db 30943 AGCTTTCAGAAAGTATTTACCATTTTATAAATCTTAAAAACATGAAAGTATATATAT 31002
Qy 5207 TTTAGAGCTCTCTATTTTCCATAGTATTCAGATTTGCAATGATCAAGTATTACTCTT 5266
Db 31003 TTTACAATAGACATATATC----- 31023
Qy 5267 TCTTTTTTTTTTTTTTTTTTTTTTTTGGATGAGTGTGTTGGTCTGTTGGCCAGCTGG 5326
Db 31024 TATTAGCTGGTGCTTTTTTTTTTTTTTCGAGATGAAGAGTCTGCTCTGTCCACCGGCTGG 31083
Qy 5327 AGTGAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCTCTGGGTTTCAACCAAGCTG 5386
Db 31084 AGTACAGTGTGGGATCTCGGCTCACTCAACCTCCACCCCGGTTTCAACGGATTCTC 31143
Qy 5387 TGGCTCAGCCCTCCCGGTTAGATGGGATTACAGGCGCCCAACCACTCGGCTAATGT 5446
Db 31144 CTGCTCAGCCCTCCCGAGTAGTGGGATTACAGGATGACACCAACCCGCTAATTT 31203
Qy 5447 TTGATTTTATAGATGAGTGGGTTTCAACCTATTTGGCCAGGCTGTCTCAACCTCCTGA 5506
Db 31204 TTGATTTTATAGAGCGGGGTTTGGCTAATGTTGGCCAGGCTGTCTCAACCTCCTGA 31263
Qy 5507 CCTCAG 5512
Db 31264 CCTCAG 31269

RESULT 5
AP002508
LOCUS 184672 bp DNA linear HTG 13-JUN-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-873E20 map 18q12, WORKING
DRAFT SEQUENCE, 33 unordered pieces.
ACCESSION AP002508
VERSION 1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone: RP11-873E20.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 184672)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 184,672 genomic DNA of 18q12
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 184672)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission

COMMENT

Submitted (09-JUN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@psc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/,
Fax: 81-42-778-9924)
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@psc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-873E20
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 166893 bases at least Q40
Consensus quality: 174794 bases at least Q30
Consensus quality: 178974 bases at least Q20
Insert size: 181472; sum-of-contigs
Quality coverage: 4.40x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
33 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 14317 contig of 14317 bp in length
14418 26921 contig of 12504 bp in length
27022 38032 contig of 11011 bp in length
38133 48298 contig of 10166 bp in length
48399 58440 contig of 10042 bp in length
58541 67531 contig of 8991 bp in length
67632 76033 contig of 8402 bp in length
76134 83796 contig of 7663 bp in length
83897 88824 contig of 4928 bp in length
88925 96427 contig of 7503 bp in length
96528 103158 contig of 6631 bp in length
103259 107080 contig of 3822 bp in length
107181 114098 contig of 6918 bp in length
114199 120020 contig of 5822 bp in length
120121 125909 contig of 5789 bp in length
126010 132285 contig of 6276 bp in length
132386 137149 contig of 4764 bp in length
137250 141731 contig of 4482 bp in length
141832 145815 contig of 3984 bp in length
145916 149496 contig of 3581 bp in length
149597 153443 contig of 3847 bp in length
153544 156388 contig of 2845 bp in length
156489 159853 contig of 3365 bp in length
159954 162765 contig of 2812 bp in length
162866 166636 contig of 3771 bp in length
166737 169695 contig of 2959 bp in length
169796 172367 contig of 2572 bp in length
172468 174970 contig of 2503 bp in length
175071 178407 contig of 3337 bp in length
178508 181049 contig of 2542 bp in length
181150 182378 contig of 1229 bp in length
182479 183536 contig of 1058 bp in length
183637 184672 contig of 1036 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 14317: contig of 14317 bp in length

repeat_region	4565..4707	/rpt_family="FRAM"	QY	3530	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCAGCCTGTAAATCCAG	3589
repeat_region	complement(4708..5016)	/rpt_family="AluY"	Db	62777	GGGGATCTTTTAAANATGTGATCGCGCGGTGGCTGGCTAGACCTGTACTCCAG	62718
repeat_region	5017..5062	/rpt_family="FRAM"	QY	3590	CACCTCTGGAGGCTGAGGTGGTGGATCGCTTTGAGCCCTGGAGTTCAAGACCAGCCTGAG	3649
repeat_region	5068..5201	/rpt_family="FLAM_C"	Db	62717	CACCTTTGGAGGCTGAGCGGGGAGATCACTTGAAGTCAAGAGTTCCAGACCAGCCTGCG	62658
repeat_region	5245..5444	/rpt_family="L1MC4"	QY	3650	CAACATGGCAAAACCCCTGTTCT-----ATAACAAAAATAGCGCGCATGTTGGCATG	3703
repeat_region	5477..5767	/rpt_family="AluJb"	Db	62657	CAATATGGCAAAACCCATCTCTACTAAAAACACCAAAATTAAGCTGGCGTATGGTGGG	62598
repeat_region	5769..5808	/rpt_family="AluJb"	QY	3704	TGCCTGTGGTCCCGAGCTA-CTAGGGGGCTGAGGAGAGAAATCTTTGGAGCCCGAGAGT	3762
repeat_region	5823..6485	/rpt_family="TA)n"	Db	62597	TGCCTGTGGTCTCAGCTACCTGGGAAGTTGAGCCATGAGGAATGATTGAACCCAGGAGT	62538
repeat_region	complement(6495..6787)	/rpt_family="L1MB3A"	QY	3763	CAAGCTGCACCTGAGCAGTGTGGGCCACTGCACCTCCAGCTGAGGTGACAGGACCAGAC	3822
repeat_region	complement(6789..6914)	/rpt_family="AluJb"	Db	62537	TAGGCTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC	62478
repeat_region	complement(7100..7282)	/rpt_family="MIR3"	QY	3823	CTTGCTCAAAAAATTAAGAGAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAA	3882
repeat_region	complement(7424..7587)	/rpt_family="FAM"	Db	62477	TCTGTCTCAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	62418
repeat_region	7646..7695	/rpt_family="(CA)n"	QY	3883	TTGCTCTAGATGAGCTACTTATAGTGGTGTATTTTGGTATTTTAACTTTTAAAGTCAGG	3942
repeat_region	complement(7713..8044)	/rpt_family="MLT1K"	Db	62417	T-----CACTCCAGCCATTCGAAATTCGCCAGGTGATGCTTAACATGCAGTAAGCGTGG	62362
repeat_region	8500..8803	/rpt_family="AluY"	QY	3943	GTCGTGCACCTGCACATTAATTAATAATCAATTCCTCAATGTATATCCACACAAAGAC	4002
repeat_region	9439..9743	/rpt_family="L1MB3"	Db	62361	AAGCATCGCACTATGCTAAGGTCTCTGGGCCGAGACCGCTGCATAAATAACCTGGGAAC	62302
repeat_region	complement(9831..10142)	/rpt_family="AluSx"	QY	4003	TGGTACGTGAATGTTTCATAGTACCTTTTATCACAACCCCAAACTAGAGACTATCCCAA	4062
repeat_region	10143..10172	/rpt_family="AT-rich"	Db	62301	AAGATACAGACACCCAGATCTCTTTATTTCAAAATCTCTGAGTTGGTCT-----CCAGG	62247
repeat_region	10519..10697	/rpt_family="MIR"	QY	4063	TATCCATCAACAAGTGAACAAATAAACAATAATGTCTATATCCATGCATGAATGAATACC	4122
repeat_region	complement(10740..10867)	/rpt_family="MER5B"	Db	62246	TATCTACCTTTTTCACAGGCTTTCAGTAATCTTGGAGAGCAGCTGCGGTGCTACTACT	62187
repeat_region	10878..11176	/rpt_family="AluSc"	QY	4123	CCTGAGTACAAAGAAAGCACTCTTGGGAGTAATCCCAAGTCATGACCTTAATGA	4182
repeat_region	11234..11584	/rpt_family="FAM"	Db	62186	CATCCAGCATAACATAAGGCAATGTTTGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGG	62127
repeat_region	11585..11717	/rpt_family="AluSq/x"	QY	4183	AAGATCAGACATGAAGGAGGAGATAATGTATGCCATAGCAATCTAGAAAT---GAA	4239
repeat_region	11721..12039	/rpt_family="AluY"	Db	62126	GACTGGGTGAATTTGGAGACTTCCATAGGCCCTCCTTATTCCTATTTGGGTTTCACTGTGG	62067
repeat_region	complement(12724..13037)	/rpt_family="AluJo"	QY	4240	AGTAACCTTATAGTTACAGAAAGCAATCAGGCGAGGCATAGAGGCTCACACCTGTAATCC	4299
repeat_region	complement(13756..14055)	/rpt_family="AluSp"	Db	62066	GGGGATGTTCATTCGAACTATTAGTTAGCGCGGCATGTTGGCTCAGCCCTGTAATCC	62007
repeat_region	complement(14362..14412)	/rpt_family="FLAM_A"	QY	4300	CAGCACTTTGAGAGGCC-ACGTGGGAAGTTGCTAGAACTCAGGAGTTCAAGACCAGCCT	4358
repeat_region	complement(14413..14703)	/rpt_family="AluJb"	Db	62006	TAGCACATTTGGAGGCCAAGCAGGTGATCACTTACCTGAGTCAAGAGTTTCGAGACCAGCT	61947
repeat_region	complement(14704..14797)	/rpt_family="FLAM_A"	QY	4359	GGCAACACAGTGAACCTCCATTCACACAAAAATGGGAAAAAAGCAAGCAATCAGTG	4418
repeat_region	14801..14998	/rpt_family="L1MC4a"	Db	61946	TGCCAACATGGCAAAACCCCATCTCTACTAAAAATACAAAAATAGCCAGGATGGTGTG	61887
repeat_region	complement(15118..15351)	/rpt_family="MER106A"	QY	4419	GTGTGCTGTGGGAGGGGAA---GGACTGCAAGAGGGAAGAGCTCTGGTGGGTGAG	4475
repeat_region	15396..15690	/rpt_family="AluSg"	Db	61886	GGGCACCTGTATCCACGCTACTTGGAGGCTGAGCGGGAGAAATCACTGGAACCCGGAG	61827
repeat_region	15701..15982	/rpt_family="AluSg"	QY	4476	GGTGTGATTCAGGTTCTGTATCCGACTGTGGGTAGCAGTTTGGGTTGTTTACATCCAAA	4535
repeat_region	16311..17324	/rpt_family="L2"	Db	61826	GCTGGAAGTTGCACTGAGCGGAGATCACACCACTTTTCACTCCAGCCTTGGTGAAGAGCAA	61767
repeat_region			QY	4536	AATATTCGTAGAAATTCATCTTTAAATGGGTGGAGTTTACTGTATGTAATATACCTC	4595
repeat_region			Db	61766	GACCCCGCTCAAAAAAAGAAAAAAGAAAAAAGAACTGTTTAAGTTAACTATGCTATGTA	61707
repeat_region			QY	4596	AATGTAAGAAAAAATATGTGTAGAAAAAGTTTCAATTCTCTTGGCAGCAACGTTATTTC	4655

Db	61706	GACCAAAATGATCTGGCAATTGATATTTTGGATGATTTTAATCTAATACCACTCACTAG	61647
QY	4656	AAATTCCTGAGCCCTTTACTTCGC AAAATCTCTGCACTTCTGCCCGTACCATTAGGTGA	4715
Db	61646	CAGACAGTTGTTCCCTGTAGACAATCGCTTTTGTAGATTTTACGTGAGCAAGATCA	61587
QY	4716	CAGCACTAGCTCCACAATTTGGATAAATGCATTTCTTGAAAAAGACTAGGACAAATCCA	4775
Db	61586	GAGTTTAAGGTTTACAAGTCACCGCAACAAGTAATAGGATGACAGTAGGTTCAGATCCCA	61527
QY	4776	GCATCAGCTTGCTTTTATATCAACCAAGCTGTACAGCTTGTGTTGCTGTCTGCAGCTG	4835
Db	61526	-----TAAAGAGAGCGGAGGTGGAGCTTGGAACTCACCATTAGGCTTAAG	61483
QY	4836	CAATGGGGACTCTGATTTCTTTAAGGAAACTTGGGTTACCAGAGATTTTCCACAAATGC	4895
Db	61482	TATCTTGATCTCTTAACCTAGTGGGAACACATAGCAAGTATTTAGATTTTGCATTA	61423
QY	4896	TATCAAAATAGTCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAAACAAGAGG	4955
Db	61422	TTTCCCTACTTATATGAATACTTAATTTCCCTTTAATCTTTTGTAGATTGCAAAAGTA	61363
QY	4956	AGGAGAAATCAGTCAATATGTGGGAACACATAGCAAGTATTTAGATTTTGCATAG	5015
Db	61362	ATATGAGCTCAATGTAGCTAATGAACA--ATGCCAGGATATTTTCTCTCACC GCCCC	61305
QY	5016	TTAAAAAGCAGCAGAGTACAAATCACACATGCAATCAGTATATCCAAATCATGTAA	5075
Db	61304	CTCCAGCCGCCACTCTCCGCCAGGTGACCAATATTACAGCAATCTCATTTCAATTTA	61245
QY	5076	TATGTGCCCTGTAGAAGACTAGAGGAATAAACA----CAAGAATCTTAACAGTCAATGTC	5131
Db	61244	GTTAGAAATGAGTTGTTCTACACAATGAATTTGTCCAAGATCAAAACTGTTCTTTC	61185
QY	5132	ATTAGACACTAGTCTAATTTATTTATTAGACACTATGATATTGAGATTTAAAAAATC	5191
Db	61184	ACTTATGTTTGTAAATACATGATCAGCAATAACTAGCAATAAATTTCTCCAGT---	61128
QY	5192	TTTAATATTTTAAAAATTTAGAGCTCTCTATTTTCCATAGTATTTCAAGTTTGACAATCA	5251
Db	61127	TTTTCTCAGATTAATTTCCATCCCGAGTTTACCCCTAATTCCTGAAGTCTGCCCCAGG	61068
QY	5252	TCAAATATPACTTCTTCTT--TTTTTTTTTTTTTTTTTTTTTTTGGAGTGGAGTTTGGT	5309
Db	61067	CAATGGTTGCTCACAAGTCAGGGTTATTATTATTTTTTTTTTTTGGAGCAAGTTTCACT	61008
QY	5310	CTTGTGCCCCATGCTGGAGTGAATGGCATGAYCATAGCTCACTGCAACCTCCACTCCT	5369
Db	61007	CTTGTGCCCCAGGCTGGAGTGAATGGCGCGATCTTGTCTACCAACAACCTCCACTCCC	60948
QY	5370	GGGTTCAAGCAAGCTGTGCGCTCAGC--CTCCCGGTTAGATGGGATTACAGCGGCCAC	5427
Db	60947	AGGTTCAAGTATTTCTCGCTCAGCTCTCCCAAGTAGCTGGGATCACAGCGGTGTC	60888
QY	5428	CACCACACTCGCTAATGTTGTATTTTGTAGAGATGGGTTTACCATGTTGGCCAG	5487
Db	60887	CACCATGCTGCTGCTAAT--TTTGTATTTTGTAGAGATGGGTTTCTCCATGTTGTCTAG	60829
QY	5488	GCTGCTCAAACTCTGACCTCAG	5512
Db	60828	GGTGCTCAAACTCTGACCTCAG	60804

RESULT 7
AC005291/c AC005291 Homo sapiens chromosome 17, clone hRPK.401_O_9, complete sequence.
DEFINITION AC005291 HTG.
AC005291.1 GI:3402737
SOURCE Homo sapiens.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 198582)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPK.401_O_9
Unpublished
2 (bases 1 to 198582)
Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatlin,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,K., Gage,D., Garadyna,S., Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Melchrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J., Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (17-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 198582)
Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,W., Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,K., Gage,D., Garadyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczyk,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Melchrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (07-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 7, 1998 this sequence version replaced gi:3399679.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.

Location/Qualifiers
1..198582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17"
/clone="hRPK.401_O_9"
/clone_lib="RPCI-11 human BAC library"
59..160
/rpt_family="LIMC5"
162..429
/rpt_family="AluSg"
complement(578..1122)
/rpt_family="MER41B"
1123..1427
/rpt_family="AluY"
complement(1428..1523)
/rpt_family="MER41B"
2090..2402
/rpt_family="AluSx"
2414..2715
/rpt_family="AluJo"
complement(2843..2927)
/rpt_family="MER57-internal"

repeat_region	3021..3315	/rpt_family="AluSp"
repeat_region	complement(4012..4318)	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(4485..4623)	
repeat_region	/rpt_family="MER5B"	
repeat_region	complement(4726..5032)	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(5033..5095)	
repeat_region	/rpt_family="MER5B"	
repeat_region	complement(5678..5998)	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(6440..6559)	
repeat_region	/rpt_family="FLAM_A"	
repeat_region	complement(6564..6866)	
repeat_region	/rpt_family="AluSc"	
repeat_region	complement(6872..7224)	
repeat_region	/rpt_family="MLTf"	
repeat_region	complement(7228..7386)	
repeat_region	/rpt_family="AluSp"	
repeat_region	complement(7409..7515)	
repeat_region	/rpt_family="MLTf"	
repeat_region	7543..7624	
repeat_region	/rpt_family="MER91A"	
repeat_region	complement(7677..7966)	
repeat_region	/rpt_family="AluSc"	
repeat_region	8535..8626	
repeat_region	/rpt_family="purine-rich"	
repeat_region	8627..8749	
repeat_region	/rpt_family="GAAA)n"	
repeat_region	8850..9244	
repeat_region	/rpt_family="MSTD"	
repeat_region	complement(9535..9818)	
repeat_region	/rpt_family="AluJo"	
repeat_region	9926..10212	
repeat_region	/rpt_family="AluX"	
repeat_region	10526..10823	
repeat_region	/rpt_family="AluSp"	
repeat_region	complement(11051..11102)	
repeat_region	/rpt_family="MADE1"	
repeat_region	11409..11719	
repeat_region	/rpt_family="AluSp"	
repeat_region	complement(11896..12214)	
repeat_region	/rpt_family="MLTfD"	
repeat_region	complement(12221..12247)	
repeat_region	/rpt_family="(CA)n"	
repeat_region	complement(12248..12515)	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(12516..12739)	
repeat_region	/rpt_family="MLTfD"	
repeat_region	12918..13226	
repeat_region	/rpt_family="AluSc"	
repeat_region	14743..14837	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(14890..15234)	
repeat_region	/rpt_family="MLTfA1"	
unsure	15298..15317	
repeat_region	/note="Single-stranded coverage."	
repeat_region	complement(15312..15590)	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(16329..16440)	
repeat_region	/rpt_family="Charlie3"	
repeat_region	complement(16446..16560)	
repeat_region	/rpt_family="L2"	
repeat_region	16565..16875	
repeat_region	/rpt_family="AluSp"	
repeat_region	complement(17280..17592)	
repeat_region	/rpt_family="AluSg"	
repeat_region	17597..17708	
repeat_region	/rpt_family="L2"	
repeat_region	17782..17906	
repeat_region	/rpt_family="MER96"	
repeat_region	complement(18625..18911)	
repeat_region	/rpt_family="AluX"	
repeat_region	18915..18949	
repeat_region	/rpt_family="L2"	
repeat_region	18950..19106	
repeat_region	/rpt_family="THE1B"	
repeat_region	complement(19107..19377)	
repeat_region	/rpt_family="AluY"	
repeat_region	19378..19589	
repeat_region	/rpt_family="THE1B"	
repeat_region	19590..19760	
repeat_region	/rpt_family="L2"	
repeat_region	19956..20415	
repeat_region	/rpt_family="MLTfG"	
repeat_region	20450..20581	
repeat_region	/rpt_family="FLAM_C"	
repeat_region	21565..21745	
repeat_region	/rpt_family="MIR"	
repeat_region	21746..21807	
repeat_region	/rpt_family="MIR"	
repeat_region	21889..22006	
repeat_region	/rpt_family="AluX"	
repeat_region	22007..22055	
repeat_region	/rpt_family="(CA)n"	
repeat_region	22056..22230	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(22333..22557)	
repeat_region	/rpt_family="MLTfJ"	
repeat_region	complement(22715..22824)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(22825..23109)	
repeat_region	/rpt_family="AluSp"	
repeat_region	complement(23110..23192)	
repeat_region	/rpt_family="MIR"	
repeat_region	24077..24285	
repeat_region	/rpt_family="AluY"	
repeat_region	complement(24558..24863)	
repeat_region	/rpt_family="AluX"	
repeat_region	complement(25044..25171)	
repeat_region	/rpt_family="MLTfA1"	
repeat_region	25172..25469	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(25470..25682)	
repeat_region	/rpt_family="MLTfA1"	
repeat_region	complement(25893..26185)	
repeat_region	/rpt_family="MLTfE"	
repeat_region	26194..26267	
repeat_region	/rpt_family="purine-rich"	
Query Match 4.4%; Score 241.8; DB 9; Length 198582;		
Best Local Similarity 50.0%; Pred. No. 5.1e-37;		
Matches 1003; Conservative 1; Mismatches 948; Indels 53; Gaps 14;		
Qy	3530	GGGAGACATTTAAAGTGAACACAGACAGCAGGCTGGTGGCTCACGCCCTGTAATCCAG 3589
Db	5058	GGGGGATCTTTTAAAAATGTGATCGCGCGGGTGGCTGGCTCAGACCTGTACTCCAG 4999
Qy	3590	CACCTCTGGAGCTGAGCTGGTGGATCGCTTGAGCCCTGGAGTTCAGACCACCCCTGAG 3649
Db	4998	CACCTTTGGAGCTGAGCGGGCAGATCACTCAAGTCAGGATTCACGACCCAGCTGGC 4939
Qy	3650	CAACATGGCAAAACCCCTGTTTCT-----ATACAAAAAATAGCCGGCATGGTGGCATG 3703
Db	4938	CAATATGGCAAAACCCATCTCTACTTAAACACCAAAATAGCCTGGCTGATGGTGGG 4879
Qy	3704	TGCCTGTGTCACGCTA-CTAGGGGCTGAGCAGGAGAACTCTTTGGAGCCAGGAGGT 3762
Db	4878	TGCCCTGTGCTCAGCTACCTGGGAAGTTGAGCCATGAGGAATGATTGAACACGAGGAGT 4819
Qy	3763	CAAGGTGCACTGAGCAGTCTTGGCCCACTGCACCTCCAGCTGGGTGACAGGACAGAC 3822
Db	4818	TAAGGTGCACTGAGTGTGTAACCTTCAGCTTCAGCTGCATGATGATGATGATGATGAT 4759
Qy	3823	CTTGCCTCAAAAAAATAAGAGAAAAATTAATAATAAATGAAACAACTACAAAGACTG 3882

[illegible]

Db	3703	TTTCCTTACTTATTATGAATACTTAAATTTCCCTTTTAACTCTTTTATCTTTTATGATGTCAGAAAGTA	3646
QY	4956	AGGAGAAATCAGTCATTATGTATGTGGGAAACAACATAGCAAGATATTTAGATCATTTTGTACTAG	5015
Db	3643	ATATGAGCTCATTGTAGCTAATGAAACA--ATGCCAGGATATTTTCTCTCACCCGCCCC	3586
QY	5016	TTAAAAAGCAGCAGAGATACAAATCACACATGCAATCAGTATTAATCCAAATCATGTAA	5075
Db	3585	CTCACGCCAGCCCTCTTCCGCCAGGTGACCAATATTAACAGGCATCTCATTTTCATTTTA	3526
QY	5076	TATGTGCTGTAGAAACAGTATAGAGTAATAACA- ---CAAGAACTTTAACAGTCATTGTC	5131
Db	3525	GTTAGAAAATCAGTGTCTTACACAAATGAATTTGTCGAAGATCAAAACTGTCCTTTTC	3466
QY	5132	ATTAGACACTAAGTCTTAATTATTATTATAGACACTATGATATTTGAGATTTTAAAAATC	5191
Db	3465	ACTTATGTTGTATATACATGATCAGCAATAACTAGCAATAATTTCTCCAGT---AGCAGC	3409
QY	5192	TTTTAAATTTTAAAAATTTAGAGCTCTCTATTTTCCATAGTATTCAGTTTTCACAATCA	5251
Db	3408	TTTTCTTCAGAAATTATTTCATCCCAAGTTTACCCCTAATTCCTGAAGTTCTGCCAGGG	3349
QY	5252	TCAAGTATTACTCTTTCTT---TTTTTTTTTTTTTTTTTTTTTTTGTAGATGGATTTTGTG	5309
Db	3348	CAAAATGCTGCTCACAAAGTCAGGTTATTATTATTTTTTTTTTTTGTAGAGCAAGTTTCACT	3289
QY	5310	CTTGTTGCCCATGCTGGAGTGGAAATGGCATGATCATGCTCACTGCAAGCTCACCTCTCT	5369
Db	3288	CTTGTTGCCCAGGCTGGAGTGCAATGGCGGATCTTTGTCTACACACAAGCTCCACCTCC	3229
QY	5370	GGGTTCAGCAAGCTCTCGCCCTCAGG--CTCCCGGGTAGATGGGATTACAGCGCCAC	5427
Db	3228	AGGTTCAAGTATCTCCTCGCTCAGCCCTCCCAAGTAGCTGGGATCACAGCGTGTGC	3169
QY	5428	CACCACTCGGCTAATGTTTGTATTATTTAGTAGAGATGGGTTTCCACCATTGTGCCAC	5487
Db	3168	CACCATGCCCTGGCTAAT- TTGTGATTTTGTAGTAGAGATGGGTTTCTCCATGTTGTGCAG	3110
QY	5488	GCTGGTCTCAAACTCTCGCTCAG	5512
Db	3109	GCTGGTCTCAAACTCTCGCTCAG	3085
RESULT 8	AC104520/c		
LOCUS	AC104520	31632 bp	DNA linear PRI 01-MAY-2002
DEFINITION	Homo sapiens chromosome 19 clone LLNLF-137B9, complete sequence.		
ACCESSION	AC104520		
VERSION	AC104520.2	GI:20376977	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 31632)		
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 31632)		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Submitted (13-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell		
REFERENCE	Drive, Walnut Creek, CA 94598, USA		
AUTHORS	3 (bases 1 to 31632)		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Submitted (01-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell		
REFERENCE	Drive, Walnut Creek, CA 94598, USA		
AUTHORS	On May 1, 2002 this sequence version replaced gi:17646891.		
TITLE	Draft Sequence Produced by DOE Joint Genome Institute		
JOURNAL	www.jgi.doe.gov		
COMMENT	Finishing Completed at Stanford Human Genome Center		
	www.sbgc.stanford.edu		

mRNA	/gene="ICBP90"	/note="Homo sapiens transcription factor ICBP90"
	<3375. .3521	
	/gene="ICBP90"	/product="ICBP90 transcription factor mRNA"
CDS	/note="This accession contains nt 2234-2382 of ICBP90 mRNA CDS AF129507."	
	<3375. .3521	
	/gene="ICBP90"	/note="ICBP90; putative transcriptional regulator of topoisomerase IIalpha expression"
repeat_region	/codon_start=1	
	/product="ICBP90 amino acids 746. .793"	
	/db_xref="GI:7577568"	
misc_feature	/translation="DCLDRSFRAQVFCSPACRYDLGRSYAMQVNPLQTLVNLQIFPCYNGNR"	
	complement(3656. .3678)	
	/rpt_family="AT-rich"	
misc_feature	complement(4013. .4372)	
	/note="BLASTN similarity to AA551654 nf98a11.s1 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:927932 3' (362. .3); match 1.00; database searched: est"	
	complement(join(4546. .4594,4617. .4869))	
misc_feature	/note="BLASTN similarity to AA282079 zt04g05.s1 NCI_CGAP_CGB1 Homo sapiens cDNA clone IMAGE:712184 3' (324. .276,253. .1); match 1.00; database searched: est"	
	complement(4959. .5010)	
	/rpt_family="MIR"	
repeat_region	complement(5087. .5276)	
	/rpt_family="MERS8A"	
	5303. .5548	
repeat_region	/rpt_family="AluJo"	
	complement(5646. .5949)	
	/rpt_family="AluJo"	
repeat_region	5951. .6041	
	/rpt_family="L1MB6"	
	complement(6056. .6357)	
repeat_region	/rpt_family="AluY"	
	6359. .6406	
	/rpt_family="L1MB8"	
repeat_region	6675. .6976	
	/rpt_family="AlusX"	
	complement(6982. .7283)	
repeat_region	/rpt_family="AluJo"	
	complement(7308. .7370)	
	/rpt_family="L1MB6"	
repeat_region	complement(7372. .7669)	
	/rpt_family="AlusX"	
	complement(7673. .8012)	
repeat_region	/rpt_family="L1MB7"	
	complement(8022. .8310)	
	/rpt_family="AlusX"	
repeat_region	complement(8320. .8603)	
	/rpt_family="AluJb"	
	8801. .8886	
misc_feature	/note="Predicted exon, program: groll2exons_human_1.3, frame: 1, quality: good, score: 62.000"	
	complement(9306. .9593)	
	/rpt_family="AluJo"	
repeat_region	complement(9618. .9903)	
	/rpt_family="AluJb"	
	complement(10229. .10489)	
repeat_region	/rpt_family="AlusX"	
	complement(10828. .11060)	
	/rpt_family="MIR"	
repeat_region	complement(11984. .12273)	
	/rpt_family="GC-rich"	
	complement(join(12530. .12561,13493. .13850))	
misc_feature	/note="BLASTN similarity to AA992428 ot53h05.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1620537 3' (402. .371,370. .13); match 0.98-1.00; database searched: est"	

repeat_region	14997..15085	/rpt_family="MIR"	
repeat_region	complement(15172..15334)		
misc_feature	/rpt_family="MER20"		
	complement(15508..15565)		
misc_feature	/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 99.000"		
	complement(15833..15893)		
repeat_region	/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: good, score: 73.000"		
repeat_region	complement(15908..16206)		
repeat_region	/rpt_family="AluSx"		
repeat_region	16462..16547		
repeat_region	/rpt_family="L1MC1"		
repeat_region	16566..16832		
repeat_region	/rpt_family="AluSx"		
repeat_region	16835..17139		
repeat_region	/rpt_family="AluY"		
repeat_region	17175..17473		
repeat_region	/rpt_family="AluSx"		
repeat_region	17522..17752		
repeat_region	/rpt_family="LINE2"		
repeat_region	17873..18185		
repeat_region	/rpt_family="MER58B"		
repeat_region	complement(18312..18610)		
repeat_region	/rpt_family="AluSx"		
repeat_region	18686..18985		
misc_feature	/rpt_family="AluSq"		
	19022..19098		
repeat_region	/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: excellent, score: 87.000"		
	complement(19726..19838)		
misc_feature	/rpt_family="(GAA)n"		
	19516..20071		
misc_feature	/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: good, score: 68.000"		
	20325..20696		
repeat_region	/note="BLASTN similarity to D44766 HUMSUPY197 Human brain cDNA Homo sapiens cDNA clone MF15-S (1..372); match 0.98; database searched: est"		
repeat_region	20943..21244		
repeat_region	/rpt_family="AluSx"		
repeat_region	21742..22017		
repeat_region	/rpt_family="AluJo"		
repeat_region	complement(22476..22538)		
repeat_region	/rpt_family="L1MB7"		
repeat_region	22539..22840		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(22841..23134)		
repeat_region	/rpt_family="L1MB8"		
repeat_region	complement(23140..23445)		
repeat_region	/rpt_family="AluSx"		
repeat_region	complement(23447..23548)		
repeat_region	/rpt_family="L1M4"		
repeat_region	24728..25023		
repeat_region	/rpt_family="AluSx"		
repeat_region	25024..25189		
repeat_region	/rpt_family="AluSg/x"		
repeat_region	complement(25236..25281)		
repeat_region	/rpt_family="Ar-rich"		
repeat_region	complement(25330..25628)		
repeat_region	/rpt_family="AluSg"		
repeat_region	complement(25652..25847)		
repeat_region	/rpt_family="MER33"		
repeat_region	26791..26909		
Query Match 4.4%; Score 241.4; DB 9; Length 41407;			
Best Local Similarity 49.4%; Pred. No. 7.2e-37;			
Matches 982; Conservative 1; Mismatches 957; Indels 47; Gaps 12;			
Qy	3539	TTTAAAGTGAACAGACAGCCAGGTGTGGTCTCAGGCTCAAGCCCTGTAATCCAGCACTCTGGG	3598
Db	39024	TTAAAGAGAAAAAGTAGTTCGGTGCAGTAGCTCAGGCTATAATCCCAACGCTTTGGG	38965

Qy	3599	AGGCTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTCGAAGACCACCTGAGCAACATGGC	3658
Db	38964	AGGCTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTCGAAGACCACCTGAGCAACATGGC	38905
Qy	3659	AAAACCTGTTTCT-----ATACAAAAATTAGCCGGGATGTGGCATGTGCTGTGG	3712
Db	38904	AAAACCTGTTTCT-----ATACAAAAATTAGCCGGGATGTGGCATGTGCTGTGG	38845
Qy	3713	TCCAGCTACT-AGGGGGCTGAGCAGGAGAATCTTTGGAGCCCGAGGAGTCAAGGCTGC	3771
Db	38844	TCCAGCTACTCAGGAGTCTGAGCAGGAGAATCAGTTGA-CTTGGGAGGAGGAGGCTGC	38786
Qy	3772	ACTGAGCAGTGTTCGCCACTGCACCTCCAGCCCTGGGTGACAGCAGCAGCTTCCCTCA	3831
Db	38785	GGTGAGCTGAGATCACACCCTGCAGGCCAGCCCTGGGCGACAGAGCAGACACTGTCTAA	38726
Qy	3832	AAAAATAAGAAGAAAAATTAATAATTAATGGAACAACACTACAAGAGCTGTTCCTTAG	3891
Db	38725	AAAAATAAGAAGAGAGAGAAAAAGTAAAGTCTCATTTAGTAGGAGAATGGGTAAATAAT	38666
Qy	3892	ATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACCTTTTAAAGTCAGGCTGTGCAC	3951
Db	38665	TACACATATCTGTACCAAAATAAAGGACTGCAGAGCCACTGAGAGAAGCAAGATCTACGCG	38606
Qy	3952	CTGCACCTACATTAATAAATATCAATTTCTCAATGTATATCCACACAAGAGCTGTACGTG	4011
Db	38605	TGACACCGCGCGCAGGAGGCCCAACACCAAGAGAGGCTTTTCCAGGTCGCGAGGATG	38546
Qy	4012	AATGTTTCATAGTACCTTTTACAAAAACCCCAAGTAGAGACTATCAAAATATCCATCA	4071
Db	38545	CAGGGACATTTGTAGTTTACACAGCCCAACCACTTACATATATCCCTTTTATAAAAAACAG	38486
Qy	4072	ACAAGTCAACAAATAAACAAATGTGCTATATCATCAATGGAATACCACCTGCCAGTA	4131
Db	38485	AGTATTTAAAGGAGCGCAGCTGGGATGCTGTGTGGGACCCCAAGTCCCTCCAGCAGCT	38426
Qy	4132	CAAGGAAGAGGCTACTTTGGGGATGAATCCCAAGTCAATGACGCTTAATGAAGAGTCAG	4191
Db	38425	CAGTGGCTGCTCGGTAGAGGATGAGAGTGCCTCTCAGGAAGAGGATGCATATTCCTGG	38366
Qy	4192	ACATGAAGGAGGAGATTAATGTATGCCATAGCAAAATTTCTAGAAAATGAAGTAACCTATAG	4251
Db	38365	GGTTGAATGATCATCAAAACGATGATGGCACCCCACTCTCT-----ACCAATTCAAAT	38314
Qy	4252	TTACAGAAACAAATCAGGCGAGCATAGAGGCTCACACCTGTAATCCAGCACTTTGAG	4311
Db	38313	TAAAAATGCCAAGCTCGGCCAGGCACAGAGGATCATGCCCTGTAATCCTAGCACTTTGGG	38254
Qy	4312	AGGCC-ACGTGGGAAGATTGCTAGAACTCAGGAGTTCGAAGACCACCTGGGCAACACAGT	4370
Db	38253	AGGCCAAGGTAGGCAGATCGTTGAGGTCAGGAGTTCGAGACCAGCTGGCCCAACATGGA	38194
Qy	4371	GAAACTCCATTCTCCACAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTGTGG	4430
Db	38193	GAAACTCCGTCCTACTTAAAAACAAAAAATTAGTGTGTGGTGGCGGTGCTGTAA	38134
Qy	4431	GGAGGGGAAGGACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGTGATTCAGGT	4490
Db	38133	TCCAGCTACTCAGGAGGCTGAGCGCAGGAGAATCGCT----TGAACCCAGGATTCTTGCT	38078
Qy	4491	TCGTATCCTTGACTGTGGTAGCAGTTTGGGGTGTGTATACATCCAAAAATATTCGTAGAAAT	4550
Db	38077	CTGTTGCCAGAGTAAAGATCCATCTCAAAAAAAGAGACCTGCCTCCAAGTGGGTGAGA	38018
Qy	4551	ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATATATACCTCAATGTAAAGAAAAAT	4610
Db	38017	CCCACTCACAGATTTGGAGTCTGTGTGAGGAAAGAGGCTACAACTCATCAGTTTCA	37958
Qy	4611	AATGTGTAAAGAAAGTTTCAATTTCTCTTGGCCAGCAACAGTTTATTCAAATTCCTGAGCCT	4670
Db	37957	GTGGTGAAGGAGATGAAAAAGAGCCAGAAACACAAAGGAGGAGGTGACAGGGGAGCTG	37898

```
QY 4671 TTACTTCGCAAAATCTCTGCACCTTCTGCCCGTACCATTAGGTGACAGACACTAGCTCCAC 4730
Db 37897 GTGTCCTCCCAAGCCCTG--CTCCGAGACAGTGGGTCTCAAAATGTTTGTATGATC 37840
QY 4731 AAATTGGATAAATGCATTTCTGAAAGACTAGGACAAATCCAGGCATCACTTGTGCT 4790
Db 37839 ACAGGTCCFAAGAGGCTTTGATGTGTTCACTTCACTGAATGTGAGGAGGAATGAGCAA 37780
QY 4791 TTCATATCAACACCGCTGTACAGCTTGTGTTGCTGTCTGCAGCTGCAATGGGCACTCTTG 4850
Db 37779 AGAAGAGGCCACCATTCAGGCAACAACACATCAGCATCTTCAAAACCCAGCCATTGA 37720
QY 4851 ATTTCCTTTAAGAAACTTGGGTACAGAGTATTTCCCAAAATGCTATTCAATATGATGC 4910
Db 37719 ACTGATTAAGTAAATCCTGATTGCTTAAGCACAGTTCAAAAGAGTAAAGCACAGCCTTGC 37660
QY 4911 TTATGAT--ATGCAAGACACTGTGCTAGGAGCCAGAAACAAGAGGAGGAGGAATCACT 4968
Db 37659 TTCAAATGATATTTGGGATCTGAAGTCGATTTCTGAGAAGTTTAAAGATGCCCTATAAGACCGA 37600
QY 4969 CATTATGTGGGAACAACATPAGCAAGATATTTAGATCATTTTGACTAGTTTAAAA--AAGCAG 5027
Db 37599 CAGCACTACTAAGAAAATCTTCCAAAATATAGTCTTAACAATCATGAAAGAAATTA 37540
QY 5028 CAGAGTACAAATCACATGCAATCAATCAGTATTAATCCAAATCATGTAATATGTCGTGA 5087
Db 37539 AATATTACATAGAAAATTTCACTTAATGCAAAAAGAAACATCAAAAGAAAGAAATAAGG 37480
QY 5088 GAAAGACTAGAGGAATAAACAACAAGAACTTAACAGTCAATCTGCTATTAGACACTAAGTCT 5147
Db 37479 ACAAATATGAATGATTAACATCCAAATGCGAGGTACAATCCAATCATATATCAATACGGC 37420
QY 5148 AATTATATTTATGACACTATGATATTTGAGATTTTAAAAAATCTT--TAATATTTTAAA 5205
Db 37419 ATTAATATGAATGATTAACATCCAAATGAGAGAGAAATTCAGACTGGATTTAA 37360
QY 5206 ATTTAGAGCTCTCTATTTTCCATAGTATTCAGAGTTTGACAAATGATCAAGTATTACHT 5265
Db 37359 AAACAAGATCAACTATATATATATATATCTGCTTCCAGGGGACA-----C 37310
QY 5266 TTCCTTTTTTTTTTTTTTTTTTTTTTTTGGATGAGATTTTGGTCTGTGTGCCATGCTG 5325
Db 37309 TTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGATTTCACTCTGTGTGCCAGGCTG 37250
QY 5326 GAGTGAATGGATGAYCATAGCTCACTGCAACCTCCACCTCCTGGTTCAGCAAGAACT 5385
Db 37249 GAGTGAATGGATGATCTTGCTCACTGCAATCTCCACCTCCCGGTTCAAGCATCTCT 37190
QY 5386 GTCGCTCAGCTCCCGGTAGATGGATTAAGGCGCCACACCACTCGGCTAATG 5445
Db 37189 CCGCTCAGCTCTGAGTGTGGGACTACAGGATGCGGCCACCGC-----A 37139
QY 5446 TTTGTATTTTATGATAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
Db 37138 TGTGTATTTTATGATAGACAAAGGTTTCTCCATGTTGTCAGCCTGGTCTCAAACTCCA 37079
QY 5506 ACCTCAG 5512
Db 37078 ACCTCAG 37072
```

```
RESULT 10
AL590609/c
LOCUS 171073 bp DNA linear HTG 26-SEP-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-4M23, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL590609
VERSION AL590609.14 GI:15795484
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 171073)
AUTHORS Wallis,J.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Sep 27, 2001 this sequence version replaced gi:15787779.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: b4M23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 170938 bases at least Q40
Consensus quality: 171073 bases at least Q30
Consensus quality: 171073 bases at least Q20
Insert size: 171073; sum-of-contrigs
Insert size: 180783; 1.3% error; agarose-fp
Quality coverage: 13.10x in Q20 bases; sum-of-contrigs Quality
coverage: 12.39x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..171073
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-4M23"
/clone_lib="RPCI-11.1"
1..171073
misc_feature
/note="assembly_fragment:05881
clone_end:SP6
vector_side:right
clone_end:T7
vector_side:left"
BASE COUNT 45416 a 38794 c 38902 g 47961 t
ORIGIN
Query Match 4.4%; Score 241.2; DB 2; Length 171073;
Best Local Similarity 63.8%; Pred. No. 6.8e-37;
Matches 453; Conservative 0; Mismatches 238; Indels 19; Gaps 5;
QY 3175 ATAAAGCTAAATATTATTGGATCTTTTATTAGTATTCATTTTATGTTTATGTTTGA 3234
Db 150354 AGACATTTCTGGTGTACTTTAAATATTATGTTATTTATTTATTTATTTATTTAT 150295
QY 3235 TTTTAAAGACAAATCTACCCCTGTTTACCCAGGCTGGAGTGCAGTGCATATAGC 3294
Db 150294 TTTCTTGAGATGGAGTCTGCTCTGCGCCAGGCTGAGTGCATATCTCGC 150235
QY 3295 TTTCTGCACTTTGAACTCTGGGCTCAAGCAATCTCTGCTTGGCTTCCCAAGTGT 3354
Db 150234 TCATGCAACCTTCCACCTTCAAGGTTTCAAGGATTTCTCTGCTCAGGCTCCCAAGTGC 150175
QY 3355 TGGGA-TACAGTTCATAGGACCACTGCATCTGGCCTAGGATTCATTTAGATTTAAATATGCA 3413
Db 150174 TAGGATTACAGCATGTGCCACCATGTCCTATTTTGTATTTTATGAGTGGAG 150115
QY 3414 TTTTAAATTTTAAATAATATGCGTAATTTTACCTTATGCTATGTAATGTAATA 3473
Db 150114 TTTCAACATGT----TGCCAGGCTGATTTTGAACCTCTGACCTCAG-----ATGATCC 150065
QY 3474 AATCTAGTTTCTGCTGCTAAAGCTTTTAAAGTGTCTTCCAAATAAGCTTCATGTCAGTCAGG-- 3531
```


Db 150064 ACCTGSCCTGCCCTCCAAAGTGCTGGGATTACAGGCATGACGACCTTGCCCGAGCGGCC 150005
Qy 3532 GGACACATTTAAAGTGAACACAGACAGCAGGTGGTGGCTCAGCGCTGTAAATCCCGACGA 3591
Db 150004 CTAGTCACTTTAAATCAATTCAGGTGGCCAGCTGGCTCAGCGCTGTAAATCCCGACGA 149945
Qy 3592 CTCTGGGAGGCTGAGGTGGGTGGATCGCTGAGCCCTGGAGTTCACAGACCGCTGAGCA 3651
Db 149944 CTTTGGAGGCGCAAGCGGGTGGATTGCTTGAGGTGAGGAGTTCACAGACCGCTGAGCA 149885
Qy 3652 ACATGGCAAAACCTCTTTCT-----ATACAAAAATTAGCCGGGATGGTGCATGTG 3705
Db 149884 ACATGGTGAACACCGGCTCTACTATAAAGTACAAAAGTTAGCCGAGCATGGTGTGCATG 149825
Qy 3706 CCTGTGTCAGCTACTAGGGGCTGAGGCGAGGAGATCTTTGGAGCCAGGAGGTCAA 3765
Db 149824 CCTGTATCTAGTACTTGGGAAGCTGAGCAGAGGAATTCCTTGAACCTGGGAGGCTAA 149765
Qy 3766 GGCTGCACTGAGCAGTCTTGGCCACTGCACTCCAGCGCTGGGTGACAGGACCGACTTT 3825
Db 149764 GGTTCAGTGAGCGAGATCATGCTGCACTCCAGCTGGGTGACAGCAAGACTCT 149705
Qy 3826 GCCTCAAAAAATTAAGAAAAATTTAAAAATAATGAACAACACTACAA 3875
Db 149704 GTCTCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 149655

RESULT 11
AL590728 270269 bp DNA linear HTG 26-JUL-2002
LOCUS Homo sapiens chromosome 1 clone RP11-45G17, *** SEQUENCING IN
DEFINITION PROGRESS **, 19 unordered pieces.
ACCESSION AL590728
VERSION AL590728.23 GI:22002653
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 270269)
Hall, R.
Direct Submission
Submitted (25-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On July 29, 2002 this sequence version replaced gi:21911471.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA45G17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 0% of reads
Sequencing vector: plasmid; 108752; 99% of reads
Chemistry: Dye-terminator; 4% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Consensus quality: 261413 bases at least Q40
Consensus quality: 264711 bases at least Q30
Consensus quality: 266087 bases at least Q20
Insert size: 268469; sum-of-contigs
Insert size: 176357; 3.4% error; agarose-fp
Quality coverage: 10.19x in Q20 bases; sum-of-contigs Quality
coverage: 17.81x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.
* 1 173248: contig of 173248 bp in length
* 173249 173248: gap of 100 bp
* 173349 178414: contig of 5666 bp in length
* 178415 178514: gap of 100 bp
* 178515 182428: contig of 3914 bp in length
* 182429 182528: gap of 100 bp
* 182529 205615: contig of 23087 bp in length
* 205616 205715: gap of 100 bp
* 205716 211474: contig of 5759 bp in length
* 211475 211574: gap of 100 bp
* 211575 225167: contig of 13593 bp in length
* 225168 225267: gap of 100 bp
* 225268 231960: contig of 6693 bp in length
* 231961 232060: gap of 100 bp
* 232061 234782: contig of 2722 bp in length
* 234783 234882: gap of 100 bp
* 234883 237701: contig of 2819 bp in length
* 237702 237801: gap of 100 bp
* 237802 240032: contig of 2231 bp in length
* 240033 240132: gap of 100 bp
* 240133 244125: contig of 3993 bp in length
* 244126 244225: gap of 100 bp
* 244226 248134: contig of 3909 bp in length
* 248135 248234: gap of 100 bp
* 248235 250284: contig of 2050 bp in length
* 250285 250384: gap of 100 bp
* 250385 252426: contig of 2042 bp in length
* 252427 252526: gap of 100 bp
* 252527 255520: contig of 2994 bp in length
* 255521 255620: gap of 100 bp
* 255621 262319: contig of 6699 bp in length
* 262320 262419: gap of 100 bp
* 262420 264483: contig of 2064 bp in length
* 264484 264583: gap of 100 bp
* 264584 267587: contig of 3004 bp in length
* 267588 267687: gap of 100 bp
* 267688 270269: contig of 2582 bp in length.
FEATURES
source
1..270269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-45G17"
/clone_lib="RPC1-11.1"
1..173248
/note="assembly_fragment:05185"
clone_end:SP6
vector_side:left
173349..178414
/note="assembly_fragment:05692"
fragment_chain:1
178515..182428
/note="assembly_fragment:07497"
fragment_chain:1
182529..205615
/note="assembly_fragment:06284"
fragment_chain:2
205716..211474
/note="assembly_fragment:06815"
fragment_chain:2
211375..225167
/note="assembly_fragment:06915"
fragment_chain:3
225268..231960
/note="assembly_fragment:07275"
fragment_chain:3
232061..234782
/note="assembly_fragment:01132"
234883..237701
/note="assembly_fragment:02015"
237802..240032

misc_feature	/note="assembly_fragment:02197"	240133..244125	
misc_feature	/note="assembly_fragment:02445"	244226..248134	
misc_feature	/note="assembly_fragment:03364"	248235..250284	
misc_feature	/note="assembly_fragment:03567"	250385..252426	
misc_feature	/note="assembly_fragment:03948"	252527..255520	
misc_feature	/note="assembly_fragment:04614"	255621..262319	
misc_feature	/note="assembly_fragment:05790"	262420..264483	
misc_feature	/note="assembly_fragment:06780"	264584..267587	
misc_feature	/note="assembly_fragment:07310.0"	267688..270369	
misc_feature	/note="assembly_fragment:07450"	270470..273151	
BASE COUNT	74304 a	62191 c	62082 g 69885 t 1807 others
ORIGIN	contig1		

RESULT 12	AC025620	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
-----------	----------	-------	------------	-----------	---------	----------	--------	----------	-----------	---------	-------	---------	-----------	---------

AC025620 155085 bp DNA linear HTG 03-FEB-2001
Homo sapiens clone RP11-27L6, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
AC025620
AC025620.3 GI:12658103
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155085)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-27L6
Unpublished
2 (bases 1 to 155085)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campilano,A., Casle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McGowan,P., McGuck,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 3, 2001 this sequence version replaced gi:7387339.
All repeats were identified using RepeatMasker:
Smith,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1965, Vol. 68, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1965, Vol. 68, No. 3, pp. 21-30
4. The Effect of Student Motivation on Learning	Journal of Educational Research	1965, Vol. 68, No. 4, pp. 31-40
5. The Role of the School in Society	Journal of Educational Research	1965, Vol. 68, No. 5, pp. 41-50
6. The Impact of Teacher Education on Classroom Practice	Journal of Educational Research	1965, Vol. 68, No. 6, pp. 51-60
7. The Importance of Professional Development	Journal of Educational Research	1965, Vol. 68, No. 7, pp. 61-70
8. The Effect of Curriculum Reform on Student Achievement	Journal of Educational Research	1965, Vol. 68, No. 8, pp. 71-80
9. The Role of the School in Promoting Social Change	Journal of Educational Research	1965, Vol. 68, No. 9, pp. 81-90
10. The Impact of Teacher Attitudes on Student Behavior	Journal of Educational Research	1965, Vol. 68, No. 10, pp. 91-100

source

*	152691	152790: gap of unknown length
*	152791	169160: contig of 16370 bp in length
*	169161	169260: gap of unknown length
*	169261	186107: contig of 16847 bp in length.

*	152691	152790: gap of unknown length
*	152791	169160: contig of 16370 bp in length
*	169161	169260: gap of unknown length
*	169261	186107: contig of 16847 bp in length.

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 155521)
TITLE	Sulston,J.E. and Waterston,R.
JOURNAL	Toward a complete human genome sequence
MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)
PUBMED	99063792
REFERENCE	9847074
AUTHORS	2 (bases 1 to 155521)
TITLE	Ozersky,P., Du,H. and Maupin,R.
JOURNAL	The sequence of Homo sapiens BAC clone RP11-723C11
REFERENCE	Unpublished (2001)
AUTHORS	3 (bases 1 to 155521)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (27-MAY-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	4 (bases 1 to 155521)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (25-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	5 (bases 1 to 155521)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (26-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	6 (bases 1 to 155521)
TITLE	Waterston,R.
JOURNAL	Direct Submission
COMMENT	Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Aug 25, 2001 this sequence version replaced gi:13431253. ----- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu/gsc Contact: sapiens@watson.wustl.edu ----- Summary Statistics ----- Center project name: H_NH0723C11
NOTICE:	This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.
MAPPING INFORMATION:	The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GRB/CHR7 , send mail to: egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc
SOURCE INFORMATION:	The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome

Db 15259 TCCATGCACTATGGAATACACTACAGCAATAAAACTACTAGTGGTATACAGAAAAACA 15318
QY 4950 AAGAGGAGGAGAAATCAGTCATTTATGTGGGAACAACATAGCAAGATATTTAGATCATTTT 5009
Db 15319 AACTACTGATTTATACACAACCTTGAATGGATCTCAGAGGATTTATGCTGAGTGAAAAA 15378
QY 5010 GACTAGTTAAAAAGCAGAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCA 5069
Db 15379 GACAATCTCAAAAGGTTACATATATGAATCCATTTATATAATTTCTGGAAGTACAGATAG 15438
QY 5070 TGTAAATATGCGCTGTAGAAGACTAGAGGAATAACACACAGAAATCTTACAGTCATTG 5129
Db 15439 ATGTGTGGTTAGCAGGGGGAGGACCGGGGTGTGAGAGTTATAAGGAATAGCATGA 15498
QY 5130 TCATTAGACACTAAGTCTAAATTTATTATTATAGACACTATGATATTTGAGATTTAAAAAA 5189
Db 15499 AAGACAGCTTTCTGAGAAATAGAACTGTTTCAGTGCCTTGACTGCTGGTGGTGAAGGAA 15558
QY 5190 TCT-TTAATATTTTAAAAATTTAGAGCTCTCTCTATTTTCCATAGTATTCAAGTTTGACAA 5248
Db 15559 TCTACGCATGTGATCAAAATTTGTATATAATTACAAAACATACACAAAATAAGTCTTGGGA 15618
QY 5249 TGATCAAGTATTACTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5308
Db 15619 CATTTATTCAATTCATGACAAAGAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 15678
QY 5309 TCTTGTGCGCATGCTGGAGTGGAAATGGCATGAYCATAGCTCACCTGCAACCTCCACCTCC 5368
Db 15679 TCTTGTACCAGGCTGGAGTGCAGTGGCAACATCTTAGTTCACCTGCAACCTCCGCCCTCC 15738
QY 5369 TGGGTTCAAGTCAAGTCTCGCCTCAGCCTCCCGGTAGATGGGATTACAGCGCCCAACC 5428
Db 15739 CGGGTTCAAGTCAAGTCTCGCCTCAGCCTCCCGGTAGATGGGATTACAGCGCGTGCC 15798
QY 5429 ACCACACTCGGCTAA-TGTTTGTATTTTAGTAGAGATGGGTTTTCACCATGTTGGCCAG 5487
Db 15799 ACCATACCCAGCTAAATTTTGTGTTTGTAGTAGAGCGGGTTTTCACCATGTTGGCTGG 15858
QY 5488 GCTGCTCAAACTCCTGACCTCAG 5512
Db 15859 GATGCTCAAAATTCCTGACCTCAG 15883

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 20:16:19 ; Search time 1128.86 seconds
(without alignments)
11000.051 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacataggtcca.....tcaaacctcctgacctcagag 5514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5512.8	100.0	5514	21 AAC55313	Human activation-i
2	4922.8	89.3	11204	21 AAC55339	Human activation-i
3	231.6	4.2	112460	24 ABK83567	Human CDNA differe
C 4	230.2	4.2	28818	22 AAL35901	Human musculoskele
C 5	227.8	4.1	112460	24 ABK83567	Human CDNA differe
C 6	227.6	4.1	4126	22 AAK80162	Human immune/haema
C 7	227.6	4.1	4126	22 AAK89461	Human digestive sy
C 8	226.8	4.1	32204	22 AAS39620	Genomic sequence #
C 9	226.8	4.1	32204	22 AAK89019	Human digestive sy

10	226.8	4.1	32204	22 AAK91533	Human digestive sy
11	226.8	4.1	32204	22 AAI57790	Human colorectal c
12	225.4	4.1	28588	22 AAK89418	Human digestive sy
13	225.4	4.1	28588	22 AAS31923	Human liver associ
14	225.4	4.1	28588	24 ABN90278	Human liver antige
15	221.6	4.0	10855	22 AAL05277	Human reproductive
16	221.6	4.0	10855	23 ABL98161	Human testicular a
C 17	221.4	4.0	10736	22 AAK80161	Human immune/haema
C 18	221.4	4.0	10736	22 AAK89460	Human digestive sy
C 19	221.2	4.0	32152	22 AAS39621	Genomic sequence #
C 20	221.2	4.0	32152	22 AAK89020	Human digestive sy
21	221.2	4.0	32152	22 AAK91534	Human digestive sy
22	221.2	4.0	32152	22 AAI57791	Human colorectal c
C 23	221.2	4.0	36933	22 AAK66362	Human immune/haema
C 24	218.8	4.0	32134	22 ABA15354	Human nervous syst
C 25	218.8	4.0	32134	22 ABA15813	Human nervous syst
C 26	218.8	4.0	32134	22 ABA15813	Human reproductive
C 27	217.8	3.9	23241	22 AAF97870	Human neuroblastom
C 28	217.8	3.9	23241	22 AAF97871	Human neuroblastom
29	216	3.9	121162	21 AAF66548	Human kinesin-like
30	215.8	3.9	154465	24 AAD28763	Human AKAP allelic
31	215.8	3.9	158245	24 AAD28762	Human AKAP allelic
32	215.8	3.9	161425	22 AAH02340	Human AKAP10 gene
33	215.8	3.9	162025	22 AAH02339	Human AKAP10 gene
34	215.8	3.9	162025	24 AAD28758	Human AKAP allelic
35	215.8	3.9	162025	24 AAD28759	Human AKAP allelic
C 36	215	3.9	19472	22 AAS26724	Human genomic DNA
C 37	214.4	3.9	17239	22 AAS38826	Human cardiovascular
C 38	213.8	3.9	14049	22 ABA15812	Human nervous syst
39	213.6	3.9	32195	22 AAS36105	Human cardiovascular
40	213.6	3.9	32195	22 AAS31532	Human DNA for a no
41	213.6	3.9	32195	22 AAS31538	Human DNA for a no
42	213.6	3.9	32195	23 ABK41045	Genomic DNA encodi
43	213.6	3.9	32195	24 ABQ66856	Human polynucleoti
44	213.6	3.9	32195	24 ABQ66862	Human polynucleoti
C 45	212.8	3.9	33147	22 AAK67282	Human immune/haema

ALIGNMENTS

RESULT 1

AAC55313

ID AAC55313 standard; DNA; 5514 BP.

XX AAC55313;

AC AAC55313;

DT 05-FEB-2001 (first entry)

DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;

KW immune related disease; allergy; allergic disease; antiallergic;

KW antianemic; antisthmatic; ophthalmological; anti-HIV; dermatologic;

KW gene therapy; B cell associated immune system disorder; food allergy;

KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;

KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;

KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;

KW ataxia telangiectasia; common variable immunodeficiency disorder;

KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

XX Homo sapiens.

OS Homo sapiens.

XX WO200058480-A1.

XX PD 05-OCT-2000.

XX PF 28-MAR-2000; 2000WO-JP01918.

XX PR 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

PR 27-DEC-1999; 99JP-0371382.

QY 1621 AAATCCCAAGCATTCCTAAATTAATTGCTTGAAGTCACTATGATGTGTCCATTAAGG 1680
Db 1621 AAATCCCAAGCATTCCTAAATTAATTGCTTGAAGTCACTATGATGTGTCCATTAAGG 1680
QY 1681 AGACAAATTCATTCAGCAAGTATTAATGCTTAAAGGCCCAATTTGTTAGGCAGTTAATG 1740
Db 1681 AGACAAATTCATTCAGCAAGTATTAATGCTTAAAGGCCCAATTTGTTAGGCAGTTAATG 1740
QY 1741 GCATCTTTACTATTAACTTAATCTTTCCATTTGTTCCAGAGTACCTTAACCTCTTAG 1800
Db 1741 GCATCTTTACTATTAACTTAATCTTTCCATTTGTTCCAGAGTACCTTAACCTCTTAG 1800
QY 1801 GTGTGAATTTGGTTAAGTCCCTCAATAATGCTTTTATGTGCAGTTTATGATAGGTTATTGT 1860
Db 1801 GTGTGAATTTGGTTAAGTCCCTCAATAATGCTTTTATGTGCAGTTTATGATAGGTTATTGT 1860
QY 1861 CATAGAACTTATCTTATTCCTACATTTATGATTACTATGGATGGATGATGAGATAACACCTA 1920
Db 1861 CATAGAACTTATCTTATTCCTACATTTATGATTACTATGGATGGATGATGAGATAACACCTA 1920
QY 1921 ATCCTTATACCTTACCTCAATTTAACTCCTTTATAAAGAACTTACATTTACAGATAAAGA 1980
Db 1921 ATCCTTATACCTTACCTCAATTTAACTCCTTTTATAAAGAACTTACATTTACAGATAAAGA 1980
QY 1981 TTTTTTAAAAATATATTTTTTTGTAGACAGGGTCTTAGCCAGCCGAGGCTGTCTCT 2040
Db 1981 TTTTTTAAAAATATATTTTTTTGTAGACAGGGTCTTAGCCAGCCGAGGCTGTCTCTCT 2040
QY 2041 AAGTCTGGCCCAAGCGATCTCCTGCTGGCCCTCCTAAAGTGTGGAATATAGACAT 2100
Db 2041 AAGTCTGGCCCAAGCGATCTCCTGCTGGCCCTCCTAAAGTGTGGAATATAGACAT 2100
QY 2101 GAGCCATCACATCCAAATACAGATAAAGATTTTTTAATGGAGGATTTAATGTTCTTCAG 2160
Db 2101 GAGCCATCACATCCAAATACAGATAAAGATTTTTTAATGGAGGATTTAATGTTCTTCAG 2160
QY 2161 AAAATTTCTTGGAGTCAGACAAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGA 2220
Db 2161 AAAATTTCTTGGAGTCAGACAAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGA 2220
QY 2221 AACAGTCTGAGCTATAGTCTTGTGAGGGTCCATTTGGAATACTTTGTTCAAAGTAAA 2280
Db 2221 AACAGTCTGAGCTATAGTCTTGTGAGGGTCCATTTGGAATACTTTGTTCAAAGTAAA 2280
QY 2281 ATGGAACCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAAGGAGAAAAG 2340
Db 2281 ATGGAACCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAAGGAGAAAAG 2340
QY 2341 ATGAAATTCACAGGACAGAGAGGAAATATATTAATCAATTAAGGAGCAGATCTGTAGA 2400
Db 2341 ATGAAATTCACAGGACAGAGAGGAAATATATTAATCAATTAAGGAGCAGATCTGTAGA 2400
QY 2401 GCTCATTTAGTCAGCAAAATGACCTTGTGTCAGGATTTATTTTAAACCCCTTGTTCCTGGT 2460
Db 2401 GCTCATTTAGTCAGCAAAATGACCTTGTGTCAGGATTTATTTTAAACCCCTTGTTCCTGGT 2460
QY 2461 TTGCACGCTGGGATGACGTAGGGTCTGCCTCAGGAGCAGACAGCTGTCAGAGCAGC 2520
Db 2461 TTGCACGCTGGGATGACGTAGGGTCTGCCTCAGGAGCAGACAGCTGTCAGAGCAGC 2520
QY 2521 TGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCTACTCAGGACAGAAAT 2580
Db 2521 TGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCTACTCAGGACAGAAAT 2580
QY 2581 GACGAGACAGGGAGCTGGAACAGGCCCTTAACGAGAGAGGGAGTAAATGGATCAACA 2640
Db 2581 GACGAGACAGGGAGCTGGAACAGGCCCTTAACGAGAGAGGGAGTAAATGGATCAACA 2640
QY 2641 AAGTTAACTAGCAGGTACAGGATCAGCAATTTCACTTCACTGTGCTGTAACATGTGACA 2700
Db 2641 AAGTTAACTAGCAGGTACAGGATCAGCAATTTCACTTCACTGTGCTGTAACATGTGACA 2700
QY 2701 GAAACAGTGTAGGCTTATTGTTTTCATGTAGAGTAGGACCCCAAAAAATCCACCCAAAGT 2760

Db 2701 GAAACAGTGTAGGCTTATTGTTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGT 2760
QY 2761 CCTTTATCTATGCCACATCCTCTCTATCTATATCTTCCAGGACACATTTTCTTCTCTATGA 2820
Db 2761 CCTTTATCTATGCCACATCCTCTCTATCTATATCTTCCAGGACATTTTCTTCTCTATGA 2820
QY 2821 TAAGGCTCTCTCTCTCCA 2880
Db 2821 TAAGGCTCTCTCTCTCCA 2880
QY 2881 CACAAC 2940
Db 2881 CACAAC 2940
QY 2941 CTCATCTACACAGCCCGGAGGGTAAAGTTAATAATTAAGAGGATTTATTTGGTAAGATGA 3000
Db 2941 CTCATCTACACAGCCCGGAGGGTAAAGTTAATAATTAAGAGGATTTATTTGGTAAGATGA 3000
QY 3001 TGCTTAAATCTGTTTAAACACTGGGCTCAAAGAGAGAATTTCTTTCTCTGCTACTTATTA 3060
Db 3001 TGCTTAAATCTGTTTAAACACTGGGCTCAAAGAGAGAATTTCTTTCTCTGCTACTTATTA 3060
QY 3061 AGCACCTATTATGTTGAGCTTATATATACAAAGGGTTATTTATATGCTTAATATAGTAAT 3120
Db 3061 AGCACCTATTATGTTGAGCTTATATATACAAAGGGTTATTTATATGCTTAATATAGTAAT 3120
QY 3121 AGTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3180
Db 3121 AGTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3180
QY 3181 CTAAATTAATTTGAGGATCTTTTAACTGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 3240
Db 3181 CTAAATTAATTTGAGGATCTTTTAACTGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 3240
QY 3241 AAAAGACAATCTCACCTGTTTACCAGGCTGGAGTGCAGTGGTGAATATGCTTTCTG 3300
Db 3241 AAAAGACAATCTCACCTGTTTACCAGGCTGGAGTGCAGTGGTGAATATGCTTTCTG 3300
QY 3301 CAGTCTTGAACCTCTGGGCTCAAGCAATCCTCTGCTTGGCTCCCAAGTGTGGGAT 3360
Db 3301 CAGTCTTGAACCTCTGGGCTCAAGCAATCCTCTGCTTGGCTCCCAAGTGTGGGAT 3360
QY 3361 ACAGTCATGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3420
Db 3361 ACAGTCATGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3420
QY 3421 TTTTAAAAATAATATGGCTAAATTTTACCTTATGTTAATGTTAATGTTAATGTTAATGTTAAT 3480
Db 3421 TTTTAAAAATAATATGGCTAAATTTTACCTTATGTTAATGTTAATGTTAATGTTAATGTTAAT 3480
QY 3481 TTTGCTGCTTAAAGTTTAAAGTGTCTTCCAAATGCTTCAATGCTTCAATGCTTCAATGCTTCAAT 3540
Db 3481 TTTGCTGCTTAAAGTTTAAAGTGTCTTCCAAATGCTTCAATGCTTCAATGCTTCAATGCTTCAAT 3540
QY 3541 TAAAGTGAACAGACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
Db 3541 TAAAGTGAACAGACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3600
QY 3601 GCTGAGGTGGTGGATCGCTTGGAGCCCTGGAGTTCAAGACCCAGCTGAGCAACATGGCAA 3660
Db 3601 GCTGAGGTGGTGGATCGCTTGGAGCCCTGGAGTTCAAGACCCAGCTGAGCAACATGGCAA 3660
QY 3661 AACCTGTTTCTATTAACAAAAATTTAGCCGGCATGGTGGCATGTGCTGCTGCTGCTGCTGCTGCT 3720
Db 3661 AACCTGTTTCTATTAACAAAAATTTAGCCGGCATGGTGGCATGTGCTGCTGCTGCTGCTGCTGCT 3720
QY 3721 ACTAGGGGCTGAGGACAGGAGTCTTTGGAGCCCGAGGAGTCAAGGCTGCACTGAGCAG 3780
Db 3721 ACTAGGGGCTGAGGACAGGAGTCTTTGGAGCCCGAGGAGTCAAGGCTGCACTGAGCAG 3780
QY 3781 TGCTTGGCCCACTGCACTCCAGCCTGGGTGACAGGACCCAGCTTGCCTCAAAAAATAA 3840

Db 3781 TGCTTCGGCCACTGCACCTCAGCCTGGGTGACAGGACCAGACCTTCGCTCAAAAAATAA 3840
QY 3841 GAAGAAAAATTAATAATGAACAACTACAAAGAGCTGTGTCTAGATGAGCTAC 3900
Db 3841 GAAGAAAAATTAATAATGAACAACTACAAAGAGCTGTGTCTAGATGAGCTAC 3900
QY 3901 TTAGTTAGGCTGATATTTTGGTATTAACTTTTAAAGTCAGGCTGTGCACCTGCACCTAC 3960
Db 3901 TTAGTTAGGCTGATATTTTGGTATTAACTTTTAAAGTCAGGCTGTGCACCTGCACCTAC 3960
QY 3961 ATTATTAAATATCAATTCATGATATATCCACAAAGAGCTGGTACGTGAATGTTAT 4020
Db 3961 ATTATTAAATATCAATTCATGATATATCCACAAAGAGCTGGTACGTGAATGTTAT 4020
QY 4021 AGTACCTTTATTACAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTAA 4080
Db 4021 AGTACCTTTATTACAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTAA 4080
QY 4081 CAAATAAACAATAATGCTATATCCATGCAATGGAATACCACTGCAGTACAAAGGAAG 4140
Db 4081 CAAATAAACAATAATGCTATATCCATGCAATGGAATACCACTGCAGTACAAAGGAAG 4140
QY 4141 AAGCTACTTGGGATGAATCCCAAGTATGAGCTTAAGTAAAGAGTCAAGATCAAGG 4200
Db 4141 AAGCTACTTGGGATGAATCCCAAGTATGAGCTTAAGTAAAGAGTCAAGATCAAGG 4200
QY 4201 AGGAGATAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTATAGTTACAGAA 4260
Db 4201 AGGAGATAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTATAGTTACAGAA 4260
QY 4261 GCNAATCAGGCGGAGGATAGAGCTCACACCTGTATCCAGACATTTTGGAGGCGACGT 4320
Db 4261 GCNAATCAGGCGGAGGATAGAGCTCACACCTGTATCCAGACATTTTGGAGGCGACGT 4320
QY 4321 GGGAACTAGCTAGAACTCAGGAGTTCAAGACAGCCTGGGCAACACAGTGAACCTCCAT 4380
Db 4321 GGGAACTAGCTAGAACTCAGGAGTTCAAGACAGCCTGGGCAACACAGTGAACCTCCAT 4380
QY 4381 TCTCCAAAAATGGGAAAAAAGAAAGCAAAATCAGTGTGTCTGTGGGAGGGGAAG 4440
Db 4381 TCTCCAAAAATGGGAAAAAAGAAAGCAAAATCAGTGTGTCTGTGGGAGGGGAAG 4440
QY 4441 GACTGCAAGAGGGAAGAGCTCTGTGTGGGTGAGGGTGGTGAATCAGGTTCTGTATCCT 4500
Db 4441 GACTGCAAGAGGGAAGAGCTCTGTGTGGGTGAGGGTGGTGAATCAGGTTCTGTATCCT 4500
QY 4501 GACTGTGTGAGCAGTTTGGGTGTTTACATCCAAAAATATTCGTAAGTATGCACTTTA 4560
Db 4501 GACTGTGTGAGCAGTTTGGGTGTTTACATCCAAAAATATTCGTAAGTATGCACTTTA 4560
QY 4561 AATGGTGGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAAATAATGTGTAAG 4620
Db 4561 AATGGTGGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAAATAATGTGTAAG 4620
QY 4621 AAAAGTTCAATCTCTGCCAGCAACAGTTATCAAAATCTCGAGCCCTTTACTTCGCA 4680
Db 4621 AAAAGTTCAATCTCTGCCAGCAACAGTTATCAAAATCTCGAGCCCTTTACTTCGCA 4680
QY 4681 AATTCTCTGACTTCTGCCCGTACCATTAGGTGACAGCAGTCTCCACAATTCGATA 4740
Db 4681 AATTCTCTGACTTCTGCCCGTACCATTAGGTGACAGCAGTCTCCACAATTCGATA 4740
QY 4741 AATGCAATTTCTGAAAAGACTAGGGACAAAAATCCAGGCATCACTTGTGCTTTTATATCAA 4800
Db 4741 AATGCAATTTCTGAAAAGACTAGGGACAAAAATCCAGGCATCACTTGTGCTTTATATCAA 4800
QY 4801 CCACGCTGTACAGCTTGTGTGCTGTGCTGCAGCTGCAATGGGAGCTCTTGAATTTCTTAA 4860
Db 4801 CCACGCTGTACAGCTTGTGTGCTGTGCTGCAGCTGCAATGGGAGCTCTTGAATTTCTTAA 4860
QY 4861 GGAACCTTGGGTTACCAGAGTATTTCCACAAATGCTTATCAATATGCTTATGATATG 4920
Db 4861 GGAACCTTGGGTTACCAGAGTATTTCCACAAATGCTTATCAATATGCTTATGATATG 4920

QY 4921 CAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAGGAAATCAGTCTATTATGTGGGA 4980
Db 4921 CAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAGGAAATCAGTCTATTATGTGGGA 4980
QY 4981 ACAACATAGCAAGATATTTAGATCATTTTGACCTAGTTTAAAAAGCAGCAGAGTACAAAT 5040
Db 4981 ACAACATAGCAAGATATTTAGATCATTTTGACCTAGTTTAAAAAGCAGCAGAGTACAAAT 5040
QY 5041 CACACATGCANTCAGTATATATCCAAATCATGTAATATGTCCTGTAGAAAGACTAGAGG 5100
Db 5041 CACACATGCANTCAGTATATATCCAAATCATGTAATATGTCCTGTAGAAAGACTAGAGG 5100
QY 5101 AATAAACACAAAGAACTTTAAACAGTCAATTTGACCTAGCATTAGACACTAAGTCTAATTTATT 5160
Db 5101 AATAAACACAAAGAACTTTAAACAGTCAATTTGACCTAGCATTAGACACTAAGTCTAATTTATT 5160
QY 5161 AGACACTATGATATTTGAGATTTTAAAAATCTTTTAATATTTTAAAAATTTAGAGCTCTTCT 5220
Db 5161 AGACACTATGATATTTGAGATTTTAAAAATCTTTTAATATTTTAAAAATTTAGAGCTCTTCT 5220
QY 5221 ATTTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTACTCTTTCTTTTTTTTTTT 5280
Db 5221 ATTTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTACTCTTTCTTTTTTTTTTT 5280
QY 5281 TTTTCTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTACTCTTTCTTTTTTTTTTT 5340
Db 5281 TTTTCTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTACTCTTTCTTTTTTTTTTT 5340
QY 5341 AYCATTAGCTCAGTCAACCTCCACCTCTGGGTTTCAAGCAAACTGTGCGCTCAGGCTCC 5400
Db 5341 AYCATTAGCTCAGTCAACCTCCACCTCTGGGTTTCAAGCAAACTGTGCGCTCAGGCTCC 5400
QY 5401 CGGGTAGATGGGATTACAGCGCCACCACACCTCGGCTAATGTTGTTATTTTAGTA 5460
Db 5401 CGGGTAGATGGGATTACAGCGCCACCACACCTCGGCTAATGTTGTTATTTTAGTA 5460
QY 5461 GAGATGGGTTTCCACCATGTTGCCAGGCTGTCTCAAACTCCTGACCTCAGAG 5514
Db 5461 GAGATGGGTTTCCACCATGTTGCCAGGCTGTCTCAAACTCCTGACCTCAGAG 5514

RESULT 2
AAC55339
ID AAC55339 standard; DNA: 11204 BP.

XX AAC55339;

XX 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.

DE Activation-induced cytidine deaminase; AID: cytidine deaminase;
XX Immune related disease; allergy; allergic disease; anti-allergic;
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

OS Homo sapiens.

XX WO200058480-A1.

PN 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

QY	2211	AGATTTTGAACAAGCTCTGAGCTATAGGTCCTTGTGAAGGTCCTCATTTGGAATACTTGT	2270
Db	1621	AGATTTTGAACAAGCTCTGAGCTATAGGTCCTTGTGAAGGTCCTCATTTGGAATACTTGT	1680
QY	2271	TCAAAGTAAATGGAAACCAAGCTTAAATCAGCAGTTGAATTCAGAGAAAGACAGAA	2330
Db	1681	TCAAAGTAAATGGAAACCAAGCTTAAATCAGCAGTTGAATTCAGAGAAAGACAGAA	1740
QY	2331	AGGAGAAAAGATGAAATTCACAGGACAGAGAGGAAATATATATCATTAAGGAGACAG	2390
Db	1741	AGGAGAAAAGATGAAATTCACAGGACAGAGAGGAAATATATATCATTAAGGAGACAG	1800
QY	2391	TATCTGTAGAGCTCATTTAGTGATGCAAAATGACTTGGTCAGGATTAATTTTAAACCCGT	2450
Db	1801	TATCTGTAGAGCTCATTTAGTGATGCAAAATGACTTGGTCAGGATTAATTTTAAACCCGT	1860
QY	2451	TGTTTCTGGTTGACGGCTGGGGATGACAGTAGGGTTCTGCCTCAGGAGACAGCTGT	2510
Db	1861	TGTTTCTGGTTGACGGCTGGGGATGACAGTAGGGTTCTGCCTCAGGAGACAGCTGT	1920
QY	2511	CCAGAGAGCTGTGACGCTGCAAGCTGAAACACACTCCCTCGGTAAAGTCCCTTCCTACTCA	2570
Db	1921	CCAGAGAGCTGTGACGCTGCAAGCTGAAACACACTCCCTCGGTAAAGTCCCTTCCTACTCA	1980
QY	2571	GGACAGAAATGACGAGAAACAGGAGCTGGAAACAGGCCCTTAACACAGAGAAGGGAAGTAA	2630
Db	1981	GGACAGAAATGACGAGAAACAGGAGCTGGAAACAGGCCCTTAACACAGAGAAGGGAAGTAA	2040
QY	2631	TGGATCAACAAGTTAACTAGCAGGTGAGGATCAGGCAATTCATTTCACTCTGACTGGTA	2690
Db	2041	TGGATCAACAAGTTAACTAGCAGGTGAGGATCAGGCAATTCATTTCACTCTGACTGGTA	2100
QY	2691	ACATGTGACAGAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATC	2750
Db	2101	ACATGTGACAGAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATC	2160
QY	2751	CACCCAAAAGTCTTTATCTATGCCACATCCTTCTTTATCTATATCTATCCAGGACACTTTTC	2810
Db	2161	CACCCAAAAGTCTTTATCTATGCCACATCCTTCTTTATCTATATCTATCCAGGACACTTTTC	2220
QY	2811	TTCCCTTATGATAAGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2870
Db	2221	TTCCCTTATGATAAGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2280
QY	2871	CACACACACACAAACACACACACCCGCCCAAGGTGCATGTAAAAAGATGTAGATTCT	2930
Db	2281	CACACACACACAAACACACACACCCGCCCAAGGTGCATGTAAAAAGATGTAGATTCT	2340
QY	2931	CTCTGCCCTTCTCATCTACAGCCCGGAGGGTAAGTTAATAAAGAGGGATTTATTGG	2990
Db	2341	CTCTGCCCTTCTCATCTACAGCCCGGAGGGTAAGTTAATAAAGAGGGATTTATTGG	2400
QY	2991	TAAGAGATGATGCTTAATCTGTTTAACTGGGCTCAAGAGAGAAATTTCTTTCTTCT	3050
Db	2401	TAAGAGATGATGCTTAATCTGTTTAACTGGGCTCAAGAGAGAAATTTCTTTCTTCT	2460
QY	3051	GTACTTATTAAGCAGCTTATTGTGTGAGCTTATATATACAAAGGGTTATTATATGCTA	3110
Db	2461	GTACTTATTAAGCAGCTTATTGTGTGAGCTTATATATACAAAGGGTTATTATATGCTA	2520
QY	3111	ATATAGTAATAGTAATGTGGTGTACTATGGTAATTTACCATAAAAATTAATATCCTTT	3170
Db	2521	ATATAGTAATAGTAATGTGGTGTACTATGGTAATTTACCATAAAAATTAATATCCTTT	2580
QY	3171	TAAAAAAGCTAATTAATTTAGGATCTTTTATAGTATTCATTTTATCTTTTATGTTT	3230
Db	2581	TAAAAAAGCTAATTAATTTAGGATCTTTTATAGTATTCATTTTATCTTTTATGTTT	2640
QY	3231	TTGATTTTTTAAAGACAACTCTACCCCTGTACCAGGCTGGAGTGCAGTGGTGAATCA	3290
Db	2641	TTGATTTTTTAAAGACAACTCTACCCCTGTACCAGGCTGGAGTGCAGTGGTGAATCA	2700

QY	3291	TAGCTTTCTGCAAGTCTTGAACCTCCCTGGGCTCAAGCAATCCTCCTGCTTGGCTTCCCAA	3350
Db	2701	TAGCTTTCTGCAAGTCTTGAACCTCCCTGGGCTCAAGCAATCCTCCTGCTTGGCTTCCCAA	2760
QY	3351	GTGTTGGGATACAGTCATGAGCCACTGCATCTGGCTAGGATCCATTTAGATTAAATAT	3410
Db	2761	GTGTTGGGATACAGTCATGAGCCACTGCATCTGGCTAGGATCCATTTAGATTAAATAT	2820
QY	3411	GCATTTTAAATTTTAAATTAATATGCTAAATTTTACCTTATGTATGTATGTTACTGGTA	3470
Db	2821	GCATTTTAAATTTTAAATTAATATGCTAAATTTTACCTTATGTATGTATGTTACTGGTA	2880
QY	3471	ATAAATCTAGTTTGTGCTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTCATGTACGTAG	3530
Db	2881	ATAAATCTAGTTTGTGCTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTCATGTACGTAG	2940
QY	3531	GGGACACATTTAAAGTGAACACAGACAGCCAGCTGGTGGCTACGCCCTGTATCCACG	3590
Db	2941	GGGACACATTTAAAGTGAACACAGACAGCCAGCTGGTGGCTACGCCCTGTATCCACG	3000
QY	3591	ACTCTGGGAGGCTGAGGTGGGTGATCGCTTTCAGCCCTGGAGTTTCAAGACACAGCTGAGC	3650
Db	3001	ACTCTGGGAGGCTGAGGTGGGTGATCGCTTTCAGCCCTGGAGTTTCAAGACACAGCTGAGC	3060
QY	3651	AACATGGCAAAACCCCTGTTTCTATACAAAAATTTAGCCGGGATGGTGCATGTGCTGT	3710
Db	3061	AACATGGCAAAACCCCTGTTTCTATACAAAAATTTAGCCGGGATGGTGCATGTGCTGT	3120
QY	3711	GGTCCAGCTACTAGGGGGCTGAGGAGAGAAATCTTTGGAGCCCGAGGTTCAAGGCTG	3770
Db	3121	GGTCCAGCTACTAGGGGGCTGAGGAGAGAAATCTTTGGAGCCCGAGGTTCAAGGCTG	3180
QY	3771	CACTGAGCAGTCTTTCGCGCACTGCACCTCCAGCCCTGGGTGACAGGACCAAGACTTGCCTC	3830
Db	3181	CACTGAGCAGTCTTTCGCGCACTGCACCTCCAGCCCTGGGTGACAGGACCAAGACTTGCCTC	3240
QY	3831	AAAAAAATTAAGAGAAAAATTTAAAAATTAATGGAAACAACTACAAAGAGCTGTGTGCTTA	3890
Db	3241	AAAAAAATTAAGAGAAAAATTTAAAAATTAATGGAAACAACTACAAAGAGCTGTGTGCTTA	3300
QY	3891	GATGAGCTACTTAGTTAGGCTGATATTTGGTATTTTAACTTTTAAAGTCAGGCTGTGCTA	3950
Db	3301	GATGAGCTACTTAGTTAGGCTGATATTTGGTATTTTAACTTTTAAAGTCAGGCTGTGCTA	3360
QY	3951	CCTGCACCTACATTTTAAAAATTAATTTCTCAATGTATATCCACACAAAGACTTGGTACGT	4010
Db	3361	CCTGCACCTACATTTTAAAAATTAATTTCTCAATGTATATCCACACAAAGACTTGGTACGT	3420
QY	4011	GAATGTTTATAGTACCTTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	4070
Db	3421	GAATGTTTATAGTACCTTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	3480
QY	4071	AACAAAGTGAACAAATTAACAAATGTGTATATCATGCAATGGAATACCACTTGCAGT	4130
Db	3481	AACAAAGTGAACAAATTAACAAATGTGTATATCATGCAATGGAATACCACTTGCAGT	3540
QY	4131	ACAAAAGGAGAGCTACTTGGGGATCAATCCCAAAAGTCATGCGCTTAAATGAAGAGTCA	4190
Db	3541	ACAAAAGGAGAGCTACTTGGGGATCAATCCCAAAAGTCATGCGCTTAAATGAAGAGTCA	3600
QY	4191	GACATGAAGGAGGAGATTAATGTATGCCATACCAATTTCTAGAAAAATGAAGTAACTTATA	4250
Db	3601	GACATGAAGGAGGAGATTAATGTATGCCATACCAATTTCTAGAAAAATGAAGTAACTTATA	3660
QY	4251	GTTACAGAAAGCAAAATCAGGCGAGCATAGAGGCTCACACCTGTATCCAGCACTTTTCA	4310
Db	3661	GTTACAGAAAGCAAAATCAGGCGAGCATAGAGGCTCACACCTGTATCCAGCACTTTTCA	3720
QY	4311	GAGGCCAGGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAAGCTTGGGCAACACAGT	4370
Db	3721	GAGGCCAGGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAAGCTTGGGCAACACAGT	3780
QY	4371	GAACTCCATTTCTCCCAAAAAATTTGGAAAAAAGCAAAATCACTGTTGCTGCTGCG	4430

QY	3488	CCTAAGTTTAAAGTGTCTTCCAAATAGCTTCATGTCGAGGGGAGACATTTTAAAGTG	3547	QY	4548	ATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTTATACCTCAANTGTAAGAAA	4607
Db	19843	AAATATGAATAGGAAAGCTTTTCTTAAGCCCTACTTTCCAAATCGGAAACCCAAAACAA	19784	Db	18794	AAAAAGAAAAAAGAAAAAATCTAAACTATATCCAAAATATTTAAAAAGTTTAACTGT	18735
QY	3548	AAACAGACA-----GCCAGGTGTGGTGGCTCACGCCCTGTAAATCCAGCAGCTCTGGG	3598	QY	4608	ATAATGTGTAAGAAAAAGTTTCAATTCCTTCCAGCAACCGTTATTTCAAATTCCTGAGC	4667
Db	19783	AAAGCCACACAGGCTTAGGCCAGATGTGTGGCTCATGCCGTGTAAATCCAGCAGCTTTGGG	19724	Db	18734	TTAGACTTTTGGCCACTGTGTCTATATTACTTTTGAGAAATTTAGAGAAAAAGCAATTTTT	18675
QY	3599	AGGCTGAGGTGGTGGATCGCTTTGAGCCCTGGAGTTCAAGACCACCGCTTGAGCAACATGGC	3658	QY	4668	CCCTTACTTCGCAAAATTCCTGCACCTTCCTGCCCCGTACCATTAGGTGACAGCAGCTAGCTC	4727
Db	19723	AGGCTTAAGCGAGGTGCACCTGAGATCAGAGTTGAGGACCGACCGCTGGCCAACTGGT	19664	Db	18674	TCAAATTTGGAAATTTGTATAAGCCCTTTTAACAGAGTAATTTACACTTTGTAGACATTTTTC	18615
QY	3659	AAAACCCGTGTCT-----ATACAAAAAATTTAGCCGGCATGTGGCATGTGCCCTTGG	3712	QY	4728	CACAAATTTGGATAAATGCAATTTCTGGAAAGACTAGGGACAAAA-----TCCAGGCA	4779
Db	19663	GAGACCCCATCTCTACTAAACTACAAAATTTAGTAGAGTGGTGTGGGTGCCCTGTAA	19604	Db	18614	CTAAAGAAATAATCAGAGATTCATATAAGATACATTTATAAGAATCGTAATACCAAAA	18555
QY	3713	TCCAGCTACTAGGG--GGCTGAGCGAGGAGAACTTTTGGAGCCAGGAGGTCAAGGCTGC	3771	QY	4780	TCACTTGTGCTTTTCATATCAACACGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCAAT	4839
Db	19603	TCCAGCTACTTTGGGAGGTGGGCGAGGAGAAATTCCTTGAAGCCAGGAGTCAGAGTTGC	19544	Db	18554	AAATTGAAATTCACAAAAATCTAATACTGGGAATTTGGTTACATAAATATGCCACATTCATAT	18495
QY	3772	ACTGAGCAGTGTTCGGCCACTGCACCTCCAGCTGGGTGACAGACCGACCTTGCCTCA	3831	QY	4840	GGGACTCTTGATTTCTTTAAGGAAACT--TGGGTTACCAGAGTATTTCACCAAAATGCTAT	4898
Db	19543	AATGAACCGAGGTCCGCCATTTGCATCCACCTGGGTGACAGGTGAGACTCTGTCTCT	19484	Db	18494	GATGAAAGATGATCGAGCTACTAAAATTTATGACATGCAAGAATATTGACATGGAAAAGTG	18435
QY	3832	AAAAAATAAGAAAGAAAAATTTAAAAATAAATGGAACAACTACAAAGAGCTGTGTCCCTAG	3891	QY	4899	TCAAATTTAGTGTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAAACAAGAGGAGG	4958
Db	19483	AAATAAATAAGC---TTCAATCATAGATAAATTTAAAGAAAAAATTAATACTGGA	19428	Db	18434	TCACACTATCTGTACATACAGGGTACAGAAACATATATAGTATGATACCTTTTGAAT	18375
QY	3892	ATGAGCTACTTAGTGTAGCTGTATTTTGGTATTTTAACTTTTAAAGTCAGGCTGTCTCAC	3951	QY	5019	AAAAAGCAGCAGAGTACAAAAATCACACATGCAATCAGTATTAATCCAAATCTGTAATAAT	5078
Db	19427	AAGGATATGTATTACAGATAGGACTTGAGGATTTTCTTACTTATAAAAAAGGCTATTTAT	19368	Db	18314	TCCTTACATAGTGGGCTTATGGGTAAATAATAATTTTATCTTTGGGCTCTTCTATCTTT	18255
QY	3952	CTGCACACTATTTTAAATATCAATTCATGTATATCCACACAAAGACTGGTACGTG	4011	QY	5079	GTGCCGTGTAAGAGACTAGAGGAATAACACAAGAATCTTAACAGTCAATTGTCATTAGAC	5138
Db	19367	ATACCAATAAATAAGGAAGATGTGAACACAGCAAAAATATCAAGGACATA-----AATAG	19313	Db	18254	-TCCAGTTTCTTAAAGACCTGTATCTTTTACATTTGGGGGTCAAGGCTGTGTTTT	18196
QY	4012	AATGTTCTATGATACCTTTTATTCACAAAACCCAAAGTAGAGACTATCCAAATATCCATCA	4071	QY	5139	ACTAAGCTAAATATTATTATTAGACACTATGATATTGAGATTTTAAAAAATCTTTAAATA	5198
Db	19312	AAATTCATGAGAGTGGTAAAAAAAATTCAA-----ACTCACTAATAATCAA	19266	Db	18195	ATTTTATTTATTTTATTTTACAGCATGGCCACTATAATATGATATATATSCCTGAAA	18136
QY	4072	ACAAGTGAACAAATAACAAAAATGTGCTATATCCATGCAATGGAATACACCCCTGCAGTA	4131	QY	5199	TTTTAAAAATTTAGAGCTCTCTATTTTCCATAGTATTCAGTTTGGACA-----	5247
Db	19265	AGGAGTACAATAATTGAATAGAGACCTTTTCTGTCTAGAAATTTAATAATGGAATAATCCA	19206	Db	18135	TTTGCTAAGAGAGATAAACCTGTTTTTTCTTTTGCATCTCTTTTTTGTATTTTAAATTT	18076
QY	4132	CAAAGGAAGAAGCTACTTCGGGATGAATCCCAAAGTCATGAGCTAAATGAAAGAGTCAG	4191	QY	5248	-----ATGATCAAGTATTACTCTCTTTTCTTTT	5274
Db	19205	GAATTGGGATATAGGAAGTAACACTCATATCTGATCATGGGAATATAAGTGGCAG	19146	Db	18075	TGTGGGTATATAATAGGTGCTTTTATTTTATTTTATTTTATTTTATTTTATTTATTTAT	18016
QY	4192	ACATGAAGGAGGAGATAATGTATGCCATACGAAATTTCTAGAAATGAAAGTAACCTTATAG	4251	QY	5275	TTTTTTTTTTTTTTTTTTTTCAGATGGAGTTTGGTCTTGTGGCCCATGCTGGAGTGGAAAT	5334
Db	19145	AAATCTGCAAGACACTTGAATATCTATCAAAAGCTTTTAAAGAGAA-----GCAT	19095	Db	18015	TTATTTATTTATTTTATTTTTCAGATCGAGTCTCGCTCTGTTTCCAGGTTTGAGTGCAGT	17956
QY	4252	TTACAGAAAGCAAAATCAGGCGAGGCATAGAGCTCACACCTGTATCCCGACACTTTGAG	4311	QY	5335	GGCATGAYCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAAGCTGTGCGCTCA	5394
Db	19094	TTCTAAGAAAAACCTAGGCTGGGTGCGATGCTCACACCTGTATATCCCGACACTTTGGG	19035	Db	17955	GGCGCAATCTCAGCTCACTGCAACCTCCACCTCCCTGGTTTCAAGCAATTTCTCTCGCTCA	17896
QY	4312	AGGCCACGTGG--GAAGATTGCTAGAACTCAGGAGTTCAAGACAGAGCTGGGCAACACAGT	4370	QY	5395	GCCTCCGGGTAGATGGGATTACAGGGCCCAACACACACTCGGCTAATGTTTGTATTTT	5454
Db	19034	AGCCAAAGCAAGTGGATCTCTTGAGCCAGGAGTTTGAGACCGAGCTGGCTTAACATGGC	18975	Db	17895	GCCTCCTGTGTAGCTGGGACTACAGGCACGCAACACACACAGCTAACATTTGTATTT	17836
QY	4371	GAACCTCCATTTCTCCACAAAATTTGGGAAAAAAGAAAGCAATTCAGTGGTTGTCCTGTGG	4430	QY	5455	TTAGTAGAGATGGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCTCGACCTCA	5511
Db	18974	AAAACCGTGTCTCTACTAATAAATAACAAAAATTAGCTGGGCTGGTGGCACACACCTGTAG	18915	Db	17835	TTAGTAGAGATGGGGTTTCCACCATGTTGGCCAGGCTGGTCTGGAACCTCTCGACCTCA	17779
QY	4431	GGAGGGGAA---GGACTGCAAGAGAGGAAGAGCTCTGGTGGGTGAGGGTGGTGATTCA	4487	RESULT 5			
Db	18914	TTCCAGCTACTTTGGGAGGTGAGCCAGGAGAAATTGCTGAACCCAGGAGTTGGAGGTGC	18855	ID	ABK83567/c		
QY	4488	GGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGCTGTTTACATCCAAAAAATATTCGTAGA	4547	cdna; 112460 BP.			
Db	18854	AATGAGCTGAGATCACACCACCTGCACCTCCAGCTTAGGTGACAGGGAGTCTCAGTCTCGAT	18795	XX			

Db	1279	TTGAGGCCACCCCTGACCAACATGTTGAAACCCCTGTCTCTACTAAAAATACAAAATTAG	1220	PR	23-AUG-2000;	2000US-0227009.
Oy	3687	CCGGCATGTGGCATGTGCTGTGTCCTCCAGCTACT-AGGGGCTGAGGCAGGAGATC	3745	PR	30-AUG-2000;	2000US-0228924.
Db	1219	CCACATGTGTGGTGGCTGCTGTATCCCACTACTCAGGAGGCTGAGGCAGGAGATC	1160	PR	01-SEP-2000;	2000US-0229287.
Oy	3746	TTTGAGCCCAAGGAGGCTCAAGGCTGCACCTGACAGTGTGTCGCCACTGCACTCCAGCCT	3805	PR	01-SEP-2000;	2000US-0229343.
Db	1159	ACTTGAACCCAGGAGGTGGAGTTGTAGTAAGCTGAGATTGAGCCACTGCACTCCAGCCT	1100	PR	01-SEP-2000;	2000US-0229344.
Oy	3806	GGGTGACAGGACAGACCTTGCCCTCAAAAAAATAAGAGAAAAATT	3851	PR	01-SEP-2000;	2000US-0229345.
Db	1099	GGGAACAGAAATGAGACTCCGCTCTCAAAAAAAGAAAAAAGT	1054	PR	05-SEP-2000;	2000US-0229509.
RESULT 8					05-SEP-2000;	2000US-0229513.
AAS39620/C					06-SEP-2000;	2000US-0230437.
ID	AAS39620 standard; DNA; 32204 BP.				08-SEP-2000;	2000US-0230438.
XX	AAS39620;				08-SEP-2000;	2000US-0231242.
XX	17-DEC-2001 (first entry)				08-SEP-2000;	2000US-0231243.
XX	Genomic sequence #39 encoding human colon associated polypeptide.				08-SEP-2000;	2000US-0231413.
DE	Human; colon cancer; congenital abnormality; infection; colitis;				08-SEP-2000;	2000US-0231414.
KW	inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;				08-SEP-2000;	2000US-0232080.
KW	intestinal inflammatory disorder; malabsorption syndrome; gastric;				08-SEP-2000;	2000US-0232081.
KW	sigmoid disease; antibacterial; antiviral; antiinflammatory;				12-SEP-2000;	2000US-0231968.
KW	cytostatic; ds.				14-SEP-2000;	2000US-0232397.
XX	Homo sapiens.				14-SEP-2000;	2000US-0232398.
PN	W0200155302-A2.				14-SEP-2000;	2000US-0232399.
XX	02-AUG-2001.				14-SEP-2000;	2000US-0232400.
XX	17-JAN-2001; 2001WO-US01240.				14-SEP-2000;	2000US-0232401.
PF	31-JAN-2000; 2000US-0179065.				14-SEP-2000;	2000US-0233063.
PR	04-FEB-2000; 2000US-0180628.				14-SEP-2000;	2000US-0233064.
PR	24-FEB-2000; 2000US-0184664.				14-SEP-2000;	2000US-0233065.
PR	02-MAR-2000; 2000US-0186350.				21-SEP-2000;	2000US-0234223.
PR	16-MAR-2000; 2000US-0189874.				21-SEP-2000;	2000US-0234274.
PR	17-MAR-2000; 2000US-0190076.				25-SEP-2000;	2000US-0234997.
PR	18-APR-2000; 2000US-0198123.				25-SEP-2000;	2000US-0234998.
PR	19-MAY-2000; 2000US-0205515.				26-SEP-2000;	2000US-0235484.
PR	07-JUN-2000; 2000US-0209467.				27-SEP-2000;	2000US-0235834.
PR	28-JUN-2000; 2000US-0214886.				27-SEP-2000;	2000US-0235836.
PR	30-JUN-2000; 2000US-0215135.				29-SEP-2000;	2000US-0236327.
PR	07-JUL-2000; 2000US-0216647.				29-SEP-2000;	2000US-0236367.
PR	07-JUL-2000; 2000US-0216880.				29-SEP-2000;	2000US-0236368.
PR	11-JUL-2000; 2000US-0217487.				29-SEP-2000;	2000US-0236369.
PR	11-JUL-2000; 2000US-0217496.				02-OCT-2000;	2000US-0236370.
PR	14-JUL-2000; 2000US-0218290.				02-OCT-2000;	2000US-0236802.
PR	26-JUL-2000; 2000US-0220963.				02-OCT-2000;	2000US-0237037.
PR	26-JUL-2000; 2000US-0220964.				02-OCT-2000;	2000US-0237038.
PR	14-AUG-2000; 2000US-0224518.				02-OCT-2000;	2000US-0237039.
PR	14-AUG-2000; 2000US-0224519.				13-OCT-2000;	2000US-0237040.
PR	14-AUG-2000; 2000US-0225213.				13-OCT-2000;	2000US-0237043.
PR	14-AUG-2000; 2000US-0225214.				13-OCT-2000;	2000US-0239935.
PR	14-AUG-2000; 2000US-0225266.				20-OCT-2000;	2000US-0239937.
PR	14-AUG-2000; 2000US-0225267.				20-OCT-2000;	2000US-0240960.
PR	14-AUG-2000; 2000US-0225268.				20-OCT-2000;	2000US-0241221.
PR	14-AUG-2000; 2000US-0225270.				20-OCT-2000;	2000US-0241221.
PR	14-AUG-2000; 2000US-0225271.				20-OCT-2000;	2000US-0241785.
PR	14-AUG-2000; 2000US-0225447.				20-OCT-2000;	2000US-0241786.
PR	14-AUG-2000; 2000US-0225757.				20-OCT-2000;	2000US-0241787.
PR	14-AUG-2000; 2000US-0225758.				20-OCT-2000;	2000US-0241808.
PR	14-AUG-2000; 2000US-0225759.				20-OCT-2000;	2000US-0241809.
PR	18-AUG-2000; 2000US-0226279.				01-NOV-2000;	2000US-0244617.
PR	22-AUG-2000; 2000US-0226681.				08-NOV-2000;	2000US-0246474.
PR	22-AUG-2000; 2000US-0226686.				08-NOV-2000;	2000US-0246475.
PR	22-AUG-2000; 2000US-0227182.				08-NOV-2000;	2000US-0246476.
PR					08-NOV-2000;	2000US-0246477.
PR					08-NOV-2000;	2000US-0246478.
PR					08-NOV-2000;	2000US-0246523.
PR					08-NOV-2000;	2000US-0246524.
PR					08-NOV-2000;	2000US-0246525.
PR					08-NOV-2000;	2000US-0246526.
PR					08-NOV-2000;	2000US-0246527.
PR					08-NOV-2000;	2000US-0246528.
PR					08-NOV-2000;	2000US-0246532.
PR					08-NOV-2000;	2000US-0246609.
PR					08-NOV-2000;	2000US-0246610.
PR					08-NOV-2000;	2000US-0246611.
PR					08-NOV-2000;	2000US-0246613.
PR					17-NOV-2000;	2000US-0249207.
PR					17-NOV-2000;	2000US-0249208.
PR					17-NOV-2000;	2000US-0249209.

Db	17593	ACAGAGCAAGACTCCGCTCTTAAACAAAAAATATATTTTTTAGGGATATGTGATCACAGC	17534	XX	02-AUG-2001.	
Qy	4773	CCAGGCATCACTTGTGCT-TTCATATCAACCAAGCTGTACAGCTTGTGTTGCTGTGCA	4831	XX	17-JAN-2001; 2001WO-US01324.	
Db	17533	TCACGTGAACCTTGAACCTCTTACTTCAAGCACTCCCTACACAGCCTACAGGTGTGTGCC	17474	XX	31-JAN-2000; 2000US-0179065.	
Qy	4832	GCTGCAATGGGACTCTTGATTTCTTTAAGGAACTTGGGTTACCAAGATATTTCCACAA	4891	PR	04-FEB-2000; 2000US-0180628.	
Db	17473	ACCCCTGGGCTAATTTTTTAAATTTTGTAGAGATGGGTCATGCTATGTGTTGCCAGG	17414	PR	24-FEB-2000; 2000US-0184664.	
Qy	4892	ATGCTATTCAAATAGTGTATGATGATCAAGACACTGTGCTAGGAGCCAGAAAAACAA	4951	PR	02-MAR-2000; 2000US-0186350.	
Db	17413	CTGCTCTCAAACTCCTGGCCTCAAGTGAGCCTCTGCTTGGCCTCCCTAAGTGCTGGGA	17354	PR	16-MAR-2000; 2000US-0189874.	
Qy	4952	GAGGAGAGAAATCAGTCAATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGA	5011	PR	17-MAR-2000; 2000US-0190076.	
Db	17353	TTATAGTGGGAGC-----CACTGCATTCGGTTGCATTTCT	17318	PR	18-APR-2000; 2000US-0198123.	
Qy	5012	CTAGTTAAAAAAGCAGCAGAGTACAAAATCACATGCAATCAATTAATCCAAATCATG	5071	PR	19-MAY-2000; 2000US-0205515.	
Db	17317	TTATTTAATTTACATCTTGATCTTTTCATATCAATATACATATCCTCATTTAATCTGTG	17258	PR	PR	2000US-0209467.
Qy	5072	TAAATATGCCCTGTAGAAGACTAGAGGAATAACACAAGAATCTTAAACAGTCAATTGC	5131	PR	28-JUN-2000; 2000US-0214886.	
Db	17257	CATAGAACAGCATAGTGTGGGTGTAGTTTATTTAGCTATTCTTTTTTTCACGGAAA--TT	17200	PR	30-JUN-2000; 2000US-0215135.	
Qy	5132	ATTAGACACTAAGCTCAATTAATTATTTATTTAGACACTATGATATTTGAGATTTAAAAATC	5191	PR	07-JUL-2000; 2000US-0216647.	
Db	17199	ATGATTGTTTACAGTTTTTTTGCATTATACCAAGAGTGTGACAGTGAACATTTGTGTACAC	17140	PR	PR	2000US-0216880.
Qy	5192	TTTAAATATTTAAATTTTAGAGCTCTTCTATTTTCCATAGTATTCAGTTTGACAAATGA	5251	PR	PR	2000US-0217487.
Db	17139	ACCTCTTTATACATACATATGTCTATTTATTTGAGAATFAGAGAAGTGGGATTTGCTGAGCTG	17080	PR	PR	2000US-0217496.
Qy	5252	TCAAGTATTACTCTTCTTT	5311	PR	PR	2000US-0218290.
Db	17079	AAAGATATGCATTTGTTTTTTTGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGAT	17020	PR	PR	2000US-0218290.
Qy	5312	TGTTGCCATGCTGGAGTGAATGGCATGAYCATGCTCACTCAACCTCCACCTCCTCGG	5371	PR	PR	2000US-0220964.
Db	17019	TGTTGCCAGGCTGGAGTAAATGGCGTGA-CTTGGCTCACTGAAACCTCTGCTCCCTA	16961	PR	PR	2000US-0220964.
Qy	5372	GTTCAAGCAAGCTGTGCGCTCAGCCTCCCGGCTAGATGGGATTTACAGGCGCCACACC	5431	PR	PR	2000US-0222513.
Db	16960	GTTCAAGCAATTTCTGCTCCTCAGCCTCCCAAGTAGCTGGGATTTACAGCACCCACC	16901	PR	PR	2000US-0225267.
Qy	5432	ACACTGGCTAA--TGTTTGTATTTTATGATGAGATGGGTTTCAACCATGTTGGCCAGGC	5489	PR	PR	2000US-0225268.
Db	16900	ACGCCAGCTAATTTTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT	16841	PR	PR	2000US-0225270.
Qy	5490	TGCTCTCAAACTCCTGACCTCA	5511	PR	PR	2000US-0225447.
Db	16840	TGCTCTCAAACTCCTGAACTCA	16819	PR	PR	2000US-0225447.
RESULT 9						
AAK89019/C						
ID	AAK89019 standard; DNA: 32204 BP.					
XX	AC	AAK89019;				
XX	AC	AAK89019;				
DT	05-NOV-2001 (first entry)					
XX	DE	Human digestive system antigen genomic sequence SEQ ID NO: 2595.				
XX	KW	Human; digestive system antigen; gene therapy; cancer; appendicitis;				
KW	KW	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;				
KW	KW	digestive system disorder; Meckel's diverticulum; ds.				
XX	OS	Homo sapiens.				
XX	XX	WO200155314-A2.				
PN	XX	WO200155314-A2.				

PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	29-SEP-2000;	2000US-0236371.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-024617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

DR	WPI; 2001-502630/55.
XX	
PT	polynucleotides encoding digestive system antigens, useful for
PT	diagnosing, treating, preventing and/or prognosing disorders of the
PT	digestive system, particularly cancer and cancer metastases -
XX	
PS	Disclosure; SEQ ID NO 2595; 986pp; English.
CC	The present invention provides the protein and coding sequences of a
CC	number of human digestive system antigens. These can be used in the
CC	diagnosis, treatment and prevention of digestive system disorders,
CC	including cancer, Meckel's diverticulum, bacterial or parasitic
CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC	ulcerative colitis. The present sequence is a genomic DNA fragment
CC	encoding a digestive system antigen of the invention.
XX	
SQ	Sequence 32204 BP; 9189 A; 7452 C; 7076 G; 8487 T; 0 other;
	Query Match 4.1%; Score 226.8; DB 22; Length 32204;
	Best Local Similarity 48.8%; Pred. No. 8.9e+29;
	Matches 976; Conservative 0; Mismatches 967; Indels 59; Gaps 11;
QY	3542 AAAGTGAACAGCAGCCAGCTGTGGTGTCAGCGCTGTAATCCCAGCACTCTGGGAGG 3601
DB	
	18793 AAATTAAAAAAAATCCCGAGTGTGGTGCACACTGTTAATCCCAGCACTCTGGGAGG 18734
QY	3602 CTGAGGTGGTGGATCGCTTTGAGCCCTGGAGTTCAAGACCAGCGCTCAGCAACATGGCAAA 3661
DB	
	18733 CCAAGACCCGGTGGATCACCTGAGTTCAGAGTTCAAGACCAGCCTGGCCAACATGGTGAA 18674
QY	3662 ACCCTGTTTTCT-----ATAACAAAAATTAGCCGGGATGGTGGCATGTGCCGTGTGTGCC 3715
DB	
	18673 ACCTGCTCTCTACTAAAAAATACAAAAATTAGCTGGGCGTGGTGTGGCGCATCTGTAATCC 18614
QY	3716 CAGCTACT-AGGGGCTGAGCAGGAGAATCTTTGGAGCGCCAGGAGTCAAGGCTGCACCT 3774
DB	
	18613 CAGCTACTCAGGAGGCTGAGCAGGAGAATGGCTTGAGCCAGGAGGCAGAGTTGCAGT 18554
QY	3775 GAGCAGTGTCTGGCCCACTGCACCTCCAGCCTGGGTGACAGCACAGACCTTGGCTCTMAAA 3834
DB	
	18553 AAGCTCAGATGATGCCACTGCACCTCCAGCCTGGGAAGACAGATGAGACTCTGTCTCAAAA 18494
QY	3835 AAATAGAGAGAAAAATTAANAATAAATGGAACAACACTACAAAGAGCTTCTTCTCCTAGATG 3894
DB	
	18493 ACAAAAACAAACAAACAAACAAANAATACAGATGCCAGGCCAGGCATGGTGGCTCACACC 18434
QY	3895 AGCTACTTAGTTAGGCTGA-----TATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTC 3949
DB	
	18433 TGTAATACAGCACTATGGAAGGCTGAGGCAGGAGGATTTGCTTAAAGCCCCAGAGTTCAAAG 18374
QY	3950 ACCTGCACACATTTTAAATATCAATTCCTCATGTATATCCACACACAAAGACTGGTACG 4009
DB	
	18373 CCAGACTGGGCAATGTAGTGAGACCCCCCATCTCTACAGGAACAATAAANAATACNAATAA 18314
QY	4010 TGAATGTTACATGTAACCTTTATTACAAAAACCCCAAGTAGAGACTATCCAAATATCCAT 4069
DB	
	18313 AAATGGCAGTTTCTCAGTTTGCCTGTCTGGGTTCTATGTGTGCTCACTGCTCTTTCAGAAATC 18254
QY	4070 CAACAAGTGAACAAATAAACAAAAATGTGCTATATCCATGCAATGGAATACCAACCTGCAG 4129
DB	
	18253 ACTGATGCCAGGTTTCACTCTTCAAGACTTGGNAGCAGCACCAAGGATGGGATTTTCG 18194
QY	4130 TACAAGGAAGACTACTTGGGGATGAATCCCAAGTCATGACCTAAATCAAAAGATC 4189
DB	
	18193 AGGCCAGACTCCAGGTATTTTAAATTTTTTTTTTAAATTTAAATTTATACAGGGGCCAAGC 18134
QY	4190 AGACATGAAGGAGGAGATAATGTATGCCATACGAAATTTCTAGAAATGAAAGTAACTTAT 4249
DB	
	18133 ACGTGGCTCACACCTGTAAATTCGGTACTTTGGGAGAACATGGCAAAACCCCTGCTCTCTA 18074
QY	4250 AGTTACAGAAGCAAAATCAGCGCAGGCATAGAGGCTCACACTGTGAATCCCAGCACTTTTG 4309
DB	
	18073 CTAAAAATACAAAAAATCAGGCTGGCGCGGTAGCTCAATGCTGTAAATCCAGCACTTTTG 18014

[illegible]

PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	PI		
PR	29-SEP-2000;	2000US-0236367.	XX	WPI; 2001-502630/55.	
PR	29-SEP-2000;	2000US-0236368.	DR		
PR	29-SEP-2000;	2000US-0236369.	XX		
PR	29-SEP-2000;	2000US-0236370.	PT	Polynucleotides encoding digestive system antigens, useful for	
PR	29-SEP-2000;	2000US-0236802.	PT	diagnosing, treating, preventing and/or prognosing disorders of the	
PR	02-OCT-2000;	2000US-0237037.	PT	digestive system, particularly cancer and cancer metastases -	
PR	02-OCT-2000;	2000US-0237038.	XX		
PR	02-OCT-2000;	2000US-0237039.	PS	Disclosure: SEQ ID NO 5109; 986pp; English.	
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	CC	The present invention provides the protein and coding sequences of a	
PR	13-OCT-2000;	2000US-0239937.	CC	number of human digestive system antigens. These can be used in the	
PR	20-OCT-2000;	2000US-0240960.	CC	diagnosis, treatment and prevention of digestive system disorders,	
PR	20-OCT-2000;	2000US-0241221.	CC	including cancer, Meckel's diverticulum, bacterial or parasitic	
PR	20-OCT-2000;	2000US-0241785.	CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or	
PR	20-OCT-2000;	2000US-0241786.	CC	ulcerative colitis. The present sequence is a genomic DNA fragment	
PR	20-OCT-2000;	2000US-0241787.	CC	encoding a digestive system antigen of the invention.	
PR	20-OCT-2000;	2000US-0241808.	XX		
PR	20-OCT-2000;	2000US-0241809.	SQ	Sequence 32204 BP; 8487 A; 7076 C; 7452 G; 9189 T; 0 other;	
PR	20-OCT-2000;	2000US-0241826.		Query Match 4.1%; Score 226.8; DB 22; Length 32204;	
PR	01-NOV-2000;	2000US-0244617.		Best Local Similarity 48.8%; Pred. No. 8.9e-29;	
PR	08-NOV-2000;	2000US-0246474.		Matches 976; Conservative 0; Mismatches 967; Indels 59; Gaps 11;	
PR	08-NOV-2000;	2000US-0246475.	QY	3542 AAAGTGAACAGACAGCCAGGTGTGGTGGCTCAGCCTGTAATCCAGCACTCTGGGAGG 3601	
PR	08-NOV-2000;	2000US-0246476.	Db	13412 AAATTAATAAATAATGCCAGTGTGGTGGCTCACACTGTAATCCAGCACTCTGGGAGG 13471	
PR	08-NOV-2000;	2000US-0246478.	QY	3602 CTGAGTGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCTGAGCAACATGGCAAA 3661	
PR	08-NOV-2000;	2000US-0246523.	Db	13472 CCAAGACCGGTGGATCACCTGAGGTGAGGTCAAGACCAGCCTGAGCAACATGGTGA 13531	
PR	08-NOV-2000;	2000US-0246524.	QY	3662 ACCCTGTTTCT-----ATAACAAATATTAGCGGGCATGGTGCATGTGCCTGTGCTCC 3715	
PR	08-NOV-2000;	2000US-0246526.	Db	13532 ACCTCGTCTCTACTATAAATAACAAAAATTAGCTGGCGTGTGTTGGGCATCTGTAATCC 13591	
PR	08-NOV-2000;	2000US-0246527.	QY	3716 CAGCTACT-AGGGGCTGAGCGAGGAATCTTTGGAGCCCGAGGAGTCAAGGCTGCAC 3774	
PR	08-NOV-2000;	2000US-0246528.	Db	13592 CAGCTACTCAGGAGGCTGAGCGAGGAATGGCTTTGAGCCAGGAGGAGGTTCAGT 13651	
PR	08-NOV-2000;	2000US-0246609.	QY	3775 GAGCAGTGTTCGCCCACTGCACCTCCAGCCTGGGTGACAGACGACAGACCTTCCTCAAA 3834	
PR	17-NOV-2000;	2000US-0249207.	Db	13652 AAGCTGAGATGATGCCACTGCACCTCCAGCCTCGAAGACAGAGTGTGTCTCAAAA 13711	
PR	17-NOV-2000;	2000US-0249210.	QY	3835 AAATAAGAAGAAAAATTAAAAATAATGGAACAACTACAAAGAGCTCTTCTCTCTAGATG 3894	
PR	17-NOV-2000;	2000US-0249212.			
PR	17-NOV-2000;	2000US-0249213.			

Db	13712	ACAAAAACAAACAAACAAAAAATACAGATGGCCAGCCAGGATGGTGGCTCACACC	13771
QY	3895	AGCTACTTATAGTGGCTGA-----TATTTTGGTATTTAACTTTTTTAAAGTCAGGGTCTGTC	3949
Db	13772	TGTAATACAGCACTATGGAAAGCTGAGCGCAGGAGGATTCCTTAAGCCCTAGAGTTCAGA	13831
QY	3950	ACCTGCACATACATATATAAAATATCAAAATTCCTCAATGTATATATCCACAAAAGACTGGTACG	4009
Db	13832	CCAGACTGGGCAATGTAGTGAGACCCCCCATCTCTACAGGAAACATAAAATACAATAA	13891
QY	4010	TGAACTTTCATAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCAT	4069
Db	13892	AAATGGCAGTTTCTCAGTTTGTCTGTGGTTCATGTGGTFCATCTCTTTCAGAAATC	13951
QY	4070	CAACAAGTGAACAAATAAACAAAAATGTGTATATCCATGCAATGAAATACCAACCTGCAG	4129
Db	13952	ACTGATGGCCAGTTCACATCTTCAAGACTGGAAAGCAGCAGCCCAAGGATGGGATTTTCG	14011
QY	4130	TACAAAGGAAGACTACTTTGGGGATGAAATCCCAAGTCATGACCTAAATCAAAGAGTC	4189
Db	14012	AGGCCAGAAGTCCAGGATATTTTAAATTTTTTAAATTTTAAATTTATACGGGGCCAAGC	14071
QY	4190	AGNATGAAGGAGGAGATTAATGTATGCCATAGCAAAATTCATAGAAATGAAGTAACTTAT	4249
Db	14072	ACGGTGGCTCACACCTGTAAATTCGCGTACTTTGGGAGAACATGGCAAAACCTGCTCTA	14131
QY	4250	AGTTACAGAAAGCAATAGGCGAGCATAGAGGCTCACACCTGTAATCCAGACACTTTG	4309
Db	14132	CTAAAAATACAAAAATCAGGCTGGCGGGTAGCTCATGCTCTAATCCAGACACTTTG	14191
QY	4310	AGAGGCCAGCT-GGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAGGCTGGGCAACACA	4368
Db	14192	GGAGGCCAAGGAGGTGGATCACTGAGATCAGGAGTCTGAGCCAGCCTGGCCCAAGCTG	14251
QY	4369	GTGAACTCCATTTCCACAAAAATGGGAAAAAAGCAAAATCATGTGTTGTCTGT	4428
Db	14252	GCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTTGTGTACACCTTT	14311
QY	4429	---GGGGAGGGGAAGGACTGCAAAGAGGGGAAGAAGCTCTGCTGGGGTGAGGGTGGTAT	4484
Db	14312	AATCCAGCTACTAAGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAAGCGTGGTT	14371
QY	4485	TCAGGTTCTGTATCTCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGT	4544
Db	14372	GCAGTAAGTCAAGATTGTGTGCCATTGCAATCCAGCCTGGGTGACAGCAAGACTCCGCT	14431
QY	4545	AGAATTATGCATCTTAATATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAGA	4604
Db	14432	CAAAAAATAAATAAATAAATAAATAAATAGTAAATAAATAAATAAATAAATAAATAAATAA	14491
QY	4605	AAAAATAATGTGTAGAAAGTTTCAATTCCTTCCAGCAACAGCTATTCAAAATTCCTG	4664
Db	14492	TGTGTTGGTGTCACTCTAGTCCAGTTACTCAGGAGGCTGAGGTGGGAGGATCACTG	14551
QY	4665	AGCCCTTTACTTCGCAAAATCTCTGCACATCTGCCCGGTA-----CCATTAGG	4712
Db	14552	AGCCCGGGAAGTTGAGGCTGCAGTGAGCTGTGATCGCACCATTGCATACAGCCTGGGTG	14611
QY	4713	TGACAGCATAGCTCCACAAATTTGGATAAATGCAATTCCTGGAAAAAGACTAGGGACAAAT	4772
Db	14612	ACAGAGCAAGACTCCGCTCTAAAAAATAAATAAATTTTTTTTAGGGATATGTGATCACAGC	14671
QY	4773	CGAGGCATCAGTTGTGCT--TTCAATACACAGGCTGTACAGCTGTGTGCTGCTGCA	4831
Db	14672	TCAGCTGAACCTTGAACCTCTTACTTCAAGAGAGTCCCTACACAGGCTACAGGTGTGCC	14731
QY	4832	CTGCAATGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACAGAGTATTTCCACAA	4891
Db	14732	ACCCCTGGCTAATTTTTTAAATTTTTTTGTAGAGATGGGTATGCTATGTTGCCCAGG	14791
QY	4892	ATGCTATTCAAATTAGTGTCTTATGATATGCAAGACACTGTCTAGGAGCCAGAAAAACAA	4951

D	b	14792	CTGCTCTCAAACTCCTGGCCCTCAAGTGAGCGCTCCTGCCCTTGGCCCTCCCTTAAGTCGTGGGA	14851		
Q	y	4952	GAGGAGGAGAAAATCAGTCATTTATGTGGGAACAACATAGCAAGATATTTTAGATCATTTTTGA	5011		
D	b	14852	TTATAGTGGGAGC-----CACITGCATTCGGTTGCACTTTCT	14887		
Q	y	5012	CTAGTTTAAAAAGCAGCAGAGTACAAAATACACATATGCANTCAGTATATATCCAATCATG	5071		
D	b	14888	TTATTAAATTTACATCTTTGACTCTTTTCATATCAATATACATATCCCTCATTTATAAAGCTGTG	14947		
Q	y	5072	TAAATATGTCCTGTAGAAGACTAGAGAGAATAAAACACAAGAATCTTAACAGTCATTCGTC	5131		
D	b	14948	CATAGAACACATAGTGTGGGTGTAGTTTATTTAGCTATCTTTTTTTTTCACCGAAA--TT	15005		
Q	y	5132	ATTAGACACTAAGCTCTAATTTATTTATTTATAGACACTATGATATTTTGAGATTTTAAAAAATC	5191		
D	b	15006	ATGATTCGTTTACAGTTTTTTTGCTATTTACCAGAGTGTTCACAGTGAACATTTCTTGTACAC	15065		
Q	y	5192	TTTAAATTTTAAAAATTTAGAGCTCTTCTATTTTTCATATTTTCCATATGATTTCAAAGTTTGACAATGA	5251		
D	b	15066	ACCTCTTTATACATACATATGCTATTTATTTGAGAAATAGAGAAGTGGGATTTGCTGAGCTG	15125		
Q	y	5252	TCAAGTATPACTCTTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTCT	5311		
D	b	15126	AAAGATATGCATTTGTTTTTTTTTTGGTTTTTTTTTTTTTTTTTTTGTAGATGGAGTTTCGCTCT	15185		
Q	y	5312	TGTTGGCCATGCTGGAGTGGAAATGGCATGAYCATAGCTCACATGCAACCTCCACACCTCCG	5371		
D	b	15186	TGTTGCCAGGCTGGAGTAAATGGCGTGA-CTTGGCTCACTGAACCTCTGCCCTCCCTA	15244		
Q	y	5372	GTTTCAAGCAAGCTGTCGCTCAGCCTCCC GGGTAGATGGGATTTACAGGCGCCCCACCACC	5431		
D	b	15245	GTTTCAAGCAATTTCTCTGCTCCTCAGCCTCCCAAGTAGCTGGGATTTACAAGCACCCACCACC	15304		
Q	y	5432	ACACTCGGCTAA--TGTTTTGATTTTTTAGTAGAGATGGGTTTACCATGTTGGCCAGGC	5489		
D	b	15305	ACGCCAGCTTAATTTTTTTTGTATTTTGTAGAGACGGGGTTTTACTAGTTGGCCAGGC	15364		
Q	y	5490	TGGTCTCAAACTCCTGACCTCA	5511		
D	b	15365	TGGTCTCCAACCTCTGAACCTCA	15386		
RESULT 11						
ID	AAI57790	standard; DNA; 32204 BP.				
XX	AAI57790;					
AC	AAI57790;					
DT	19-OCT-2001	(first entry)				
DE	Human colorectal cancer antigen coding sequence SEQ ID NO: 327.					
XX	Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.					
KW	Homo sapiens.					
OS	WO200155350-A1.					
PN	02-AUG-2001.					
XX	17-JAN-2001;	2001WO-US01350.				
PR	31-JAN-2000;	2000US-0179065.				
PR	04-FEB-2000;	2000US-0180628.				
PR	24-FEB-2000;	2000US-0184664.				
PR	02-MAR-2000;	2000US-0186350.				
PR	16-MAR-2000;	2000US-0189874.				
PR	17-MAR-2000;	2000US-0190076.				
PR	18-APR-2000;	2000US-0198123.				
PR	19-MAY-2000;	2000US-0205515.				
PR	07-JUN-2000;	2000US-0209467.				
PR	28-JUN-2000;	2000US-0214886.				

PR	30-JUN-2000;	2000US-0215135.	PR	20-OCT-2000;	2000US-0241787.
PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000;	2000US-0216880.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	14-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224518.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246611.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226681.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226868.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	01-DEC-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250391.
PR	12-SEP-2000;	2000US-0231968.	PR	03-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232399.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232400.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	21-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235834.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	XX	WPI; 2001-457727/49.	
PR	29-SEP-2000;	2000US-0236367.	XX		
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	PT	Isolated polypeptide for treating, preventing and/ or prognosing	
PR	29-SEP-2000;	2000US-0236370.	PT	disorders related to the colon and rectum including colorectal cancers	
PR	02-OCT-2000;	2000US-0236802.	PT	and also for testing and detection e.g. diagnosis -	
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237038.	PS	Disclosure; SEQ ID NO: 327; 522pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237039.	XX		
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	CC	The present invention provides the protein and coding sequences of a	
PR	13-OCT-2000;	2000US-0239937.	CC	number of colorectal cancer antigens. These are shown in	
PR	20-OCT-2000;	2000US-0240930.	CC	AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the	
PR	20-OCT-2000;	2000US-0240960.	CC	diagnosis, prevention and treatment of cancer of the colon and/or rectum.	
PR	20-OCT-2000;	2000US-0241221.	CC	The present sequence is a colorectal cancer antigen genomic sequence.	
PR	20-OCT-2000;	2000US-0241785.	CC	Note: The sequence data for this patent did not form part of the printed	
PR	20-OCT-2000;	2000US-0241786.	CC	specification, but was obtained in electronic format directly from WIPO	

CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 32204 BP; 8487 A; 7076 C; 7452 G; 9189 T; 0 other;

Query Match		4.1%;	Score 226.8;	DB 22;	Length 32204;
Best Local Similarity		48.8%;	Pred. No. 8.9e-29;		
Matches 976;		Conservative 0;	Mismatches 967;	Indels 59;	Gaps 11;
QY	3542	AAAGTGAACAGACAGCCAGGTGGTGGCTCAGCCCTGTAATCCAGCACTCTGGGAGG	3601		
DB	13412	AAATTAATAAATGCGAGGTGGTGGCTCAGCCCTGTAATCCAGCACTCTGGGAGG	13471		
QY	3602	CTGAGGTGGGTGGTGGCTTGGAGCCCTGGAGTTCAGAGCAGCCCTGAGCAACATGGCAAA	3661		
DB	13472	CAAAGACCGGTGGATCACTGAGGTGAGGTGCAAGACCAAGCCCTGGCCACATGGTGA	13531		
QY	3662	ACCTGTTTCT- - - - -ATACAAAAATTAGCCGGGCATGGTGGCATGTGCTGTGTC	3715		
DB	13532	ACCTGCTCTACTAAAAATACAAAAATTAGTGGCGTGGTGGGCATCTGTAAATCC	13591		
QY	3716	CAGCTACT- AGGGGGCTCAGGAGGAGATCTTTGGAGCCAGGAGGTCAAGGCTGCAC	3774		
DB	13592	CAGTACTCAGAGGCTCAGGAGGAGATGGCTTGAGCCAGGAGGAGGTGCACT	13651		
QY	3775	GAGCAGTCTTGGCCACTGCACCTCAGCCCTGGGTGACAGGACCAAGCTTGCCTCAAAA	3834		
DB	13652	AAGCTGAGATGATGCCACTGCACATCCAGCCCTGGAAGACAGAGTGAGACTGTCTCAAAA	13711		
QY	3835	AAATTAAGAAAAATTAATAATAATGAAGAACTACAAGAGCTGTTGCTTAGATG	3894		
DB	13712	ACAAAAACAACAACAAAAATAACAGATGGCCAGGCCAGGATGGTGGCTCACACC	13771		
QY	3895	AGCTACTTAGTTAGGCTCA- - - - -TATTTTGGTATTAACTTTTAAAGTCAGGGTCTGTC	3949		
DB	13772	TGTAATACAGCACTATGAAGGCTCAGCAGGAGGATGGCTTAAGCCCTAGAGTTCAAGA	13831		
QY	3950	ACCTGCACACTACATTAATAATATCAATTTCTCAATGTATATCCACAAAAAGACTGGTACG	4009		
DB	13832	CCAGACTGGGCAATGTAGTGAGACCCCCCATCTCTACAGGAACAATAAAAAACAATAA	13891		
QY	4010	TGAATGTTCAATAGTACCTTTATTACAAAAACCCAAAGTAGAGACTATCCAAATATCCAT	4069		
DB	13892	AAATGGCAGTTTCTCAGTTTGTGCTGGTGGTCTATGTGCTCACTCTCTTCAGAATC	13951		
QY	4070	CAACAAGTGAACAAATAACAAAAATGTCTATATCCATGCAATGAATACCAACCTGCAG	4129		
DB	13952	ACTGATGCCAGTTTCACTCTTCAAGACTGGAAGCAGCAGCAGCAAGGATGGGATTTTCG	14011		
QY	4130	TACAAAGGAAGAAGCTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAGAAGTC	4189		
DB	14012	AGGCCAGAATCCAGGTATTTTAATTTTAAATTTTAAATTTATACAGGGGCCAAGC	14071		
QY	4190	AGACATGAAGGAGGAGATAATGTATGCCATAGAAATTTCTAGAAATGAAGTAACTAT	4249		
DB	14072	ACGGTGGCTCACACCTGTAATTCCTGGTACTTTGGGAGAACATGGCAAAACCTGTCTCTA	14131		
QY	4250	AGTTACAAAGCAATCAGGCAGGCATAGAGCTCACCTGTAATCCAGCACTTTG	4309		
DB	14132	CTAAATATACAAAAAATCAGGCTGGGCGGTAGTCTAGCTGCTGTAATCCAGCACTTG	14191		
QY	4310	AGAGGCCACGT- GGGAGATTGCTAGAACTCAGGAGTTCAAGACCAGGCTGGGCAACACA	4368		
DB	14192	GGAGCCAAAGCAGGTGATCACCTGAGATCAGGAGTCTGAGCCAGGCTGGCCAAAGTG	14251		
QY	4369	GTGAACCTCCATTTCTCCAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTGCTGT	4428		
DB	14252	CGGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTTGTGTACACCTTT	14311		
QY	4429	- - - - -GGGAGGGGAAGGACTGCAAGAGGGAAGAGCTCTGTTGGGGTGAAGGTGGTAT	4484		
DB	14312	AATCCAGCTACTAAGAGGAGGCTGAGGCAAGAGAAATGCTTGAACCTGGGAAGCGGTGTT	14371		
QY	4485	TCAGGTTCTGTATCTGACTGTGGTAGCAGTTTGGGGTGTTTTACATCCAAAAATATTCGT	4544		

DB	14372	GCAGTAAGTCAAGATTGTGCCAATTCGAATCCAGCCTGGGTGACAGCAAGACTCCGCT	14431		
QY	4545	AGAATTATGCATCTTAATGGTGGAGTTACTGTATCTAAATTTATACCTCAATTAAGA	4604		
DB	14432	CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	14491		
QY	4605	AAAAATAATGTGAAGAAAGTTTCAATTCCTCTGCCAGCAACCTTATTCAAATTCCTG	4664		
DB	14492	TGTGTTGTTGCTCAGCTGAGTCCAGTTACTCAGGAGGCTCAGGTGGGAGATCACCTG	14551		
QY	4665	AGCCCTTACCTTTCGCAAAATCTCTGCACCTCTGCCCCGTA- - - - -CCATTAAG	4712		
DB	14552	AGCCCGGAAGTGTGAGGCTGAGCTGAGTGCACCTGATCGCACCTACAGCTCGGGTG	14611		
QY	4713	TCACAGCACTAGCTCCACAAATTTGGATAAATGCAATTTCTGGAAGAGACTAGGACAAAT	4772		
DB	14612	ACAGAGCAAGACTCCGCTTAAAAAATAAATTTTATTTTAGGATATGTATCACAGC	14671		
QY	4773	CCAGGCATCACTTGTGCT- TTTATATCAACACAGCTGTACAGCTTGTGTCTGTCA	4831		
DB	14672	TCAGTGTAACTTGAATCTTACTTCAAGCAGTCTTACCAGGCTTACAGGTGTGCTGCC	14731		
QY	4832	GCTGCAATGGGAGCTCTTGAATTTCTTAAAGAAACTTGGTTCACAGAGTATTTCCACAA	4891		
DB	14732	ACCCCTGGCTAAATTTTAAAAATTTTGTAGAGATGGGTGATGTGTGCCAGG	14791		
QY	4892	ATGCTATTCAATTAAGTCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAAACA	4951		
DB	14792	CTGCTCTCAAACTCTGCCCCCAAGTGAGCCCTCTGCCCCCTCCCTTACAGTGTGGGA	14851		
QY	4952	GAGGAGGAGAAATCAGTCATTTATGTGGGAACAACATAGCAAGATATTAGATCATTTGA	5011		
DB	14852	TTATAGTTGGGAGC- - - - -CACTGCATTCGTTGCATTTCT	14887		
QY	5012	CTAGTTAAAAAGCAGCAGAGTACAAAATCACAAATGCAATGCAATTAATCCAAATCATG	5071		
DB	14888	TTATTTAATATCATCTTGTATCTTTTCATATCAATATACATATPCTTCATTTAACTGTTG	14947		
QY	5072	TAAATATGCTGCTGTAGAAAGCTAGAGGAATAAACAAGAACTTAAACAGTCATGTC	5131		
DB	14948	CATAGACAGCAGTAGTGTGGGTGATTTATTTAGCTATTTCTTTTTCACGGA- -TT	15005		
QY	5132	ATTACAGCACTAAGTCTAATATTATTTATAGACACTATGATATTGAGATTTAAAAAATC	5191		
DB	15006	ATGATTTGTACAGTTTTTGTCTATTACCAAGAGTTTGACAGTGAACATTTCTGTACAC	15065		
QY	5192	TTTAATATTTAAAAATTTAGAGCTCTTCTATTTTCCATAGPATTTCAAGTTTGACAATGA	5251		
DB	15066	ACCTCTTTATACATACATATGTATTTTGAAGATAGAGAAGTGGGATTTGCTAGCTG	15125		
QY	5252	TCAGTATATCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT	5311		
DB	15126	AAAGATATGCAATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT	15185		
QY	5312	TGTTCCCATGCTGGAGTGGAAATGCATAGCATAGCTACATGCACTGCAACCTCCACCTCGG	5371		
DB	15186	TGTTCCCATGCTGGAGTGAATAAATGGCGTGA-CTTGCGTCACTGAACCTCTGCTCCCTA	15244		
QY	5372	GTTCAGCAAAAGCTGCTGCGCTCAGCCTCCCGGTAGTGGGATTTACAGCGCCGCCACCC	5431		
DB	15245	GTTCAGCAAAATTTCTCTGCTCAGCCTCCCAAGTAGTGGGATTTACAAGCACCCACACC	15304		
QY	5432	ACACTCGGCTAA- -TGTTTGTATTTTGTAGTAGAGATGGGTTTACCATGTTGGCAGGC	5489		
DB	15305	ACGCCAGCTAATTTTGTATTTTGTAGAGCGGGTTTTTACTAGCTTGGCGAGC	15364		
QY	5490	TGGTCTCAAACTCTGACCTCA	5511		
DB	15365	TGGTCTCAAACTCTGAACTCA	15386		

AAK89418
ID AAK89418 standard; DNA; 28588 BP.
XX
AC AAK89418;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 2994.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0228688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
DR
XX Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX Disclosure; SEQ ID NO 2994; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 28588 BP; 7983 A; 5600 C; 5729 G; 9276 T; 0 other;

Query Match 4.1%; Score 225.4; DB 22; Length 28588;
Best Local Similarity 48.3%; Pred. No. 1.5e-28;
Matches 1154; Conservative 1; Mismatches 1142; Indels 94; Gaps 15;

QY 3213 TTTATGTTTTTATGTTTTTAAAGACAATCTCACCTGTTTACCAGGCTGG 3272
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19521 TTTTATTTTATTAATTTTTTTTGAGTAGGAGTTTCTGTTGCCAGCGGG 19580
QY 3273 AGTCAGTGGTCAATAGATCTTCTGCAGCTTTGAATCCTGGGCTCAAGCAATCCTC 3332
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19581 AGTGAAGTGGCGTGATCTCAGCTCACTGCAACCTCCACCTCTGGGTTCAAGCGATTCTC 19640
QY 3333 CTGCGCTTGGCTCCCAAGTGTGGAT---ACATCATAGCCCACTGCATCTGCGCTAG 3389
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19641 CTGCGTCAAGCTCTCTGAGTAGTGGGATTACAGGTGCCCAACCCATGCCAGCTAATTT 19700
QY 3390 GATCCATTATAGATTAATATGCAATTTTAAATTTTAAATATATGG---CTAATTTT 3445
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19701 TTTATATTTTAGTAGACGGGTTTCCCGTGTGACCGGTGGTCTCAAACTCCTG 19760
QY 3446 ACCTTATTAAT-----GTGTATCTGGTAATAATCTAGTTTGCCTTAAAGTTTA 3498
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19761 ACCTCAGTGATCCACCTGCGCTTGGCTCCCAAGAGTTGGGATTACAGGCATGAGCCAC 19820
QY 3499 AAGTCTTTCCATAGCTTC-----ATGTACGTGAGGGAGACATTTAAGTGAACA 3552
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19821 CGTGCCTGGCCAAAATAATTTTCTGATTTTATAGAGAAAGTTTATAAAAATAGTTAT 19880
QY 3553 GACAGCCAGGTGTGTGCTACGCTGTAATCCAGCAGCTCTGGAGGCTGAGGTGGT 3612
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19881 GCAGCTGGCCAGGTGGCTATGCTGTAAATCCAGCACATTTGGAGGCTGAGGCAGGT 19940
QY 3613 GGATCGTTGAGCCCTGGAGTTCAGACCCAGCTTGAGCAACATGGCAAAACCTGTTCT 3672
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
19941 GGATCATTGAGCTTAGGAGTTGAGATCGGCTTGAGCAACATGGCAAAACCTGCTTT 20000
QY 3673 ATACAAA-----AATAGCCGGGATGGTGGCATGTGCTGTGTCAGCTACTA 3724
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20001 ACCAAAAATACAAACCAATTACCCAGGTGCGAGTGGCATGTGCTGTGTCAGATACTC 20060

QY 3725 GGGGCTGAGGCAGGAGAAATCTTTGGAGCCCGAGAGGTCAAGGCTGCACTGAGCAGTGCT 3784
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20061 AGAGGCTGAGGCAGGAGGATTACCTGAGCCTGGAGGCGAGGTTGTCAGTGACAT 20120
QY 3785 TCGGCCACTGCACTCCAGCCTGGGTGACAGGACACAGACTTCCCTCAAAAAAATAAGAAG 3844
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20121 CGTGCCACTGCAATCAGCCTGGGCAACAGAGTGAGATCTCTCTCAAAAAACAACAAA 20180
QY 3845 AAAAATTAATAATGTAAGAACTACAAAGAGC-----TGTGTGTC 3888
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20181 AAAACGGTTATAGCATAGCAAGGCTTTATGATATGATATAGAATAGTTTATTTAT 20240
QY 3889 TAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTAAAGCTCAGGGTCTGT 3948
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20241 TTTATGTCATCTTAATTTATATATATATTAATTTGATTAAGCCCTAGTGAATTTTAGTTTCA 20300
QY 3949 CACCTGCACTACATTTAATAATCAATCTCAATGTATATCCACACAAGACTGGTAC 4008
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20301 AGACTTAATGGTTTTTTTTTCTTTTAAATTTACATTTGAATGAGGCTTTTCATATAC 20360
QY 4009 GTGAATGTTTCATAGTAC-CTTTATTACAAAAACCCCAAGTAGAGACTATCCAAATATCC 4067
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20361 TATACCTTTTATATCAGCAATGATGGTTAATTTTAAATAGCACTTTACAGTATATACTTAC 20420
QY 4068 ATCAACAAGTGAACAAATAAACAATAATGTCTTATATCCATGCAATGGAATACCACTGTC 4127
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20421 ATTTTTCATAGATCTTAACACCATTTCTTATTACATTTCTAGTTTATCTCCAGTA 20480
QY 4128 AGTACAAGGAAGAAGC---TACTTTGGGATGAATCCCAAGTCATGACGCTAAATGAA 4183
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20481 TAGTTAAAGATTAAGCCATTGTAAGTGTGACAGGTAATAATTAAGCAATTTAGGAATAA 20540
QY 4184 AGAGTCAGACATGAAGGAGGAGATTAATGTATGCCATACCAAT-TCTAGAAATGAACT 4242
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20541 AATGCTCCACAGGGAACCTGAAACTTATTGCTTGAATGAGTGTCTTTGGTTTATAA 20600
QY 4243 AACTTATAGTTACAAAAACAATCAGGCGAGCATAGAGGCTCACACCTGTAATCCCAAG 4302
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20601 AAGTGAATGAAGTATATATCTCGAGGCATGTGAATCATTTGACCGCTTTGCACAAAG 20660
QY 4303 CACTTTGAGAGGCCACGTGGGAAAGATGCTAGAACTCAGAGGTTCAAGACCACGCTGGGC 4362
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20661 GCTAAATGAATATTAGTAGAATAATTCAGTAATAATAGGACTTGAATCTGTTTTCTC 20720
QY 4363 RACAGTGAACCTCCATCTCCACAAAAATGGGAAAAAAGAAAGCAATCAGTGGTTG 4422
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20721 TGTATGTTAAATAAATTTTCAAGAAATACAGGTTAGGCTATGAACAATTTATGATTT 20780
QY 4423 TCCTGTGGGAGGGAAGGACTGCAAGAGGGAAGA---AGCTCTGGTGGGTGAGGGTG 4479
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20781 TATTAATAATCATAGAAAGGACTACTTATCTCATATTTGGAAGTTTATATCATTTGCTG 20840
QY 4480 GTGATTCAGGTTTCTATCTCTGACTGTGGTAGCAGTTTGGGTGTTTACATCCAAATA 4539
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20841 TTTACTTAGTTTTCAGCATATTGCTTCCCTCATTTTAAAGTAAATTTCTCTAAGCATA 20900
QY 4540 TTCGTGAATATTGATCATTTAATGGTGGAGTTTACTGTATGTAAATTTATACCTCAATG 4599
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20901 TTTCTAGCTTGTCTTCTGCTTCTGATTTGATATATTTTCTCTCCTAGAACTCAAC 20960
QY 4600 TAAGAAAAAATAATGTAAGAAAAAGTTTCAATTTCTTGGCCAGCAAGCTTATCAAAAT 4659
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
20961 ATTGAGCTTTACACCTAGCTTCTGAGATCACTGTAAATGGAAGAAAGAGCCAGATAAT 21020
QY 4660 TCCTGAGCCCTTTACTTCGAAATTTCTTGCCTACT-TCTGCCCGTACCATTTAGGTGACAG 4718
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
21021 TCTTGGCCATACCTTCCCTTCAAACTTATCCAGGATGTCTTCAATTTTGCAGCTCTTGGAG 21080
QY 4719 CACTAGCTCCCAAAATTTGATTAATGTCATTTCTGGAAGAGACTAGGACAAAAA-----T 4772
DB IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
21081 CAGTTCCATCTCTTATGTTTACCTCTGCAATCGAGGCTCTGAAGACTCTACATGTCTC 21140

QY	4773	CCAGGATCAGCTTGCTTTTCATATCAACACCGCTGTACAGCTTGTTGCTGTGCAG	4832	KW	anti-infertility.
Db	21141	GAATGAAAGCATTTCAAAACGGAATGGCAGTTGCCAGTTCTCGAATCTAATCCTT	21200	XX	Homo sapiens.
QY	4833	CTGCAATGGGACTCTTGATTTCTTAAGGAACCTGGTTACCAAGATAT	4883	PN	WO200155355-A1.
Db	21201	GGGTAGAAAGTTATTTATCTCGTATGTTAATTTGATTTCTTAAGCAGATCTACTAGGA	21260	XX	02-AUG-2001.
QY	4884	TTCCCAATGCTATTCAAAATAGTCTTATGATATGAAGACACTGTGCTAGGAGCCAG	4943	PF	17-JAN-2001; 2001WO-US01351.
Db	21261	TTTCAGCAGTTATCTCGATCAGTCGTTTATGTATACATTTATTTAGTTAGTGCCTTT	21320	XX	31-JAN-2000; 2000US-0179065.
QY	4944	AAACAAGAGAGAGAGAAATCAGTCATTA	4981	PR	04-FEB-2000; 2000US-0180628.
Db	21321	CACGTATTTGTTTGTTCATTTATTTATTTAGCAGCTGCCATGTGTCAGGCTCTGTGTGGG	21380	PR	24-FEB-2000; 2000US-0184664.
QY	4982	CAACATAGCAAGATATTAGATCATTTTTCAGCTAGTTAAAGAGCAGCAGATACAAATC	5041	PR	02-MAR-2000; 2000US-0186350.
Db	21381	TCAGGTGGAAAGATAGACCTGGTCTCTACTGTGTATTCGCTCTAGTTTGTAGTGGATG	21440	PR	16-MAR-2000; 2000US-0189874.
QY	5042	ACACATGCAATCAGTATATCCAAATCATGTAAATATGTCCTGTAGAAAGACTAGAGGA	5101	PR	17-MAR-2000; 2000US-0190076.
Db	21441	CTTGGATAGTACAAATCAGTTTATGCAATTAGTTTCTCTGGTGTCTGTGATTTGA	21500	PR	18-APR-2000; 2000US-0198123.
QY	5102	ATAACACAAGATCTTAACAGTCATGTGATTAAGACACTAAGTCTAATATATATATTA	5161	PR	19-MAY-2000; 2000US-0205515.
Db	21501	AGGCTTGTGTATACCTATTCAAGATTTTAACATTAGACTTTTCAGTTGATTGCTTATA	21560	PR	07-JUN-2000; 2000US-0209467.
QY	5162	GACACTATGATTTTGAGATTTTAAATAATCTTAATATTTTAAATTTAGAGCTCTTCTA	5221	PR	30-JUN-2000; 2000US-0215135.
Db	21561	ATCTGTCTGCTGTATTCAGCATTAGTGTCACTTTTAATCTGATAGAGGTTTGTCCCA	21620	PR	07-JUL-2000; 2000US-0216647.
QY	5222	TTTTTCCATAGTATTCAGATTTGACAATGATCAAGTATTAACCTCTTCTTTTTTTTTTTT	5281	PR	07-JUL-2000; 2000US-0216880.
Db	21621	GTGAGACTGATATTAGTGTCTAAATCCCTCAGGTTCTCTCTCTCTCTTTTTTTTTT	21680	PR	11-JUL-2000; 2000US-0217487.
QY	5282	TTTTTTTTTTTGAGATGAGTTTTGGTCTTGTCGCCATGCTGGAGTGAATGCATGA	5341	PR	11-JUL-2000; 2000US-0217496.
Db	21681	TTTTTTTTTTTGAGACGAGGTTTCACCTC---TTGCCAGGCTGGAGTGCATGATGAA	21737	PR	14-JUL-2000; 2000US-0218290.
QY	5342	YCATAGCTCACATGCAACCTCCACCTCTGGGTTCAAGCAAAAGCTGTCGCTCAGCTCCC	5401	PR	26-JUL-2000; 2000US-0220964.
Db	21738	TTTCAGCTCAGTCGAACCTCTGCCCGGTTCAAGTGATTTCTCTGCTCAGGCTCCC	21797	PR	14-AUG-2000; 2000US-0224518.
QY	5402	GGGTAGATGGGATTACAGCGCCACACACATCGGCTAATGTTGTTATTTTATTTAGTAG	5461	PR	14-AUG-2000; 2000US-0225213.
Db	21798	GAGTAGCTGGGATTACAGGTGCGCCGCCACACGCGCCAGCTATTTTGTATTTTAGTAA	21857	PR	14-AUG-2000; 2000US-0225267.
QY	5462	AGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG	5512	PR	14-AUG-2000; 2000US-0225268.
Db	21858	AGACGGGGTTTCACCATGTTGGCCAGTCTGGTCTCGAACTCCTGACCTCAG	21908	PR	14-AUG-2000; 2000US-0225270.
RESULT 13					PR 14-AUG-2000; 2000US-0225447.
AAS31923					PR 14-AUG-2000; 2000US-0225757.
ID	AAS31923 standard; DNA; 28588 BP.				PR 14-AUG-2000; 2000US-0225758.
XX					PR 14-AUG-2000; 2000US-0226279.
AC					PR 22-AUG-2000; 2000US-0226681.
XX					PR 22-AUG-2000; 2000US-0226868.
DT					PR 22-AUG-2000; 2000US-0227182.
XX					PR 30-AUG-2000; 2000US-0227009.
DT					PR 01-SEP-2000; 2000US-0229287.
XX					PR 01-SEP-2000; 2000US-0229343.
DT					PR 01-SEP-2000; 2000US-0229344.
XX					PR 05-SEP-2000; 2000US-0229345.
DT					PR 05-SEP-2000; 2000US-0229509.
XX					PR 05-SEP-2000; 2000US-0229513.
DT					PR 06-SEP-2000; 2000US-0230437.
XX					PR 06-SEP-2000; 2000US-0230438.
DT					PR 08-SEP-2000; 2000US-0231242.
XX					PR 08-SEP-2000; 2000US-0231243.
DT					PR 08-SEP-2000; 2000US-0231244.
XX					PR 08-SEP-2000; 2000US-0231413.
DT					PR 08-SEP-2000; 2000US-0231414.
XX					PR 08-SEP-2000; 2000US-0232080.
DT					PR 12-SEP-2000; 2000US-0232081.
XX					PR 14-SEP-2000; 2000US-0232397.
DT					PR 14-SEP-2000; 2000US-0232398.
XX					PR 14-SEP-2000; 2000US-0232399.
DT					PR 14-SEP-2000; 2000US-0232400.
XX					PR 14-SEP-2000; 2000US-0232401.
DT					PR 14-SEP-2000; 2000US-0233063.
XX					PR 14-SEP-2000; 2000US-0233065.
DT					PR 21-SEP-2000; 2000US-0234223.
XX					PR 21-SEP-2000; 2000US-0234274.
DT					PR 25-SEP-2000; 2000US-0234997.

Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;
chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
cerebrovascular disorder; nervous system disorder; bacterial infection;
fungal infection; viral infection; ocular disorder; endocrine disorder;
gastrointestinal disorder; renal disorder; respiratory disorder;
wound healing; skin aging; organ transplantation; tissue regeneration;

PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457728/49.
XX Isolated nucleic acid molecule encoding a human liver related protein
is used in preventing, treating or ameliorating disorders of the liver
particularly cancer of the liver -
XX Claim 1; SEQ ID No 399; 526pp; English.
XX Sequences AAS31827-AAS32182 represent genomic DNA molecules, which encode
the liver associated polypeptides of the invention. Liver associated
polypeptides and their associated polynucleotides are useful in the
diagnosis, treatment and prevention of various types of disorders in e.g.
humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
pathological condition can be determined by detecting the presence or
absence of a mutation in a liver associated polynucleotide. The treatable
disorders include autoimmune diseases such as rheumatoid arthritis,
hyperproliferative disorders such as neoplasms of the breast or liver,
cardiovascular disorders such as cardiac arrest, cerebrovascular
disorders such as cerebral ischaemia, nervous system disorders such as
Alzheimer's disease, infections caused by bacteria, viruses and fungi,
ocular disorders such as corneal infection, endocrine disorders such as
premature labour and infertility, gastrointestinal disorders such as
Crohn's disease, renal disorders such as glomerulonephritis and
respiratory disorders such as asthma and pleurisy. The polypeptides can
also be used to aid wound healing, to prevent skin aging due to sunburn,
to maintain organs before transplantation, to regenerate tissues and in
chemotaxis.
XX Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Query Match 4.1%; Score 225.4; DB 22; Length 28588;
Best Local Similarity 48.3%; Pred No. 1.5e-28;
Matches 1154; Conservative 1; Mismatches 1142; Indels 94; Gaps 15;
QY 3213 TTTATGTTTTTTTATGTTTTTTCATTTTAAAGACAATCTCACCTGTACCAGGCTGG 3272
DB 19521 TTTTATGTTTTTATTAATTTTGTGAGTGGAGTTTCTACCTGTGCCAGGCGGG 19580
QY 3273 AGTCAGTGGTGCATATCATAGCTTCTGCAGTCTTGAACCTCTGGGCTCAAGCATCTCTC 3332
DB 19581 AGTGAAGTGGCGTGATCTCAGCTCACTGCAACCTCCACCTCTCTGGTTCAGGCGATTCTC 19640
QY 3333 CTGCCTTGGCCTCCCAAGTGTGGGAT---ACAGTCATGAGCCACTGCATCTGGCCTAG 3389
DB 19641 CTGCCTCAGCCTCTCTGAGTAGCTGGGATTACAGGTGCCACCACCATGCCAGCTAATTT 19700
QY 3390 GATCCATTTAGATTAATAATATGCATTTTAAATTTTAAATTAATATG---CTAATTTTT 3445
DB 19701 TTTATATTTTAGTAGACACGGGTTTCCACCTGTTGACCAGCTGGTCTCAAACTCTG 19760
QY 3446 ACCTATGTAAT-----GTGTATCTGGTAAATAAATCTAGTTTGTCCCTAAAGTTTA 3498
DB 19761 ACCTCAGGTGATCCACCTTGGCCTCCCAAGTGTGGGATTACAGGCATGAGCCAC 19820
QY 3499 AAGTGCTTTCCAATAAGCTTC-----ATGTACGTGAGGGGAGACATTTAAAGTGAACA 3552
DB 19821 CGTCCCTGGCCAAAAAATTTTCTGATTTTAGAGGAAGTGTTATATAAATAGTATT 19880
QY 3553 GACAGCCAGGTGTGGTGGCTCAGGCTGTAATCCAGCACTCTGGAGGCTGAGGTGGGT 3612
DB 19881 GCAGGCTGGGCACGGTGGCTCATGCTGTAATCCAGCACATTGGAGGCTGAGGCAGGT 19940
QY 3613 GGATCGCTTGAGCCCTGGAGTTCAAGCAGGCTGAGCAACATGCGCAAAACCCCTGTTCT 3672
DB 19941 GGATCACCTTGAGCTTAGGAGTTTGAGATCGGCCTGAGCAACATGCGCAAAACCCCTGTTCT 20000

QY	3673	ATAACAAA-----AATTAGCGGCGATGGTGGCATGTGCTGTGTGCCAGCTACTA	3724	Db	21081	CAGTTCCACTCTCTATTTGATTGTTACCTCTGCAATCGAGGCTCTGAAGACTCTACATGTCC	21140
Db	20001	ACCAAAATACAAACCAATTAGCCAGGTGCAGTGCATGTGCTGTGGTCCCGATACTC	20060	QY	4773	CCAGGCATCAGTTGCTGCTTTCATATCAACACGCTGTACACGCTGTGTTGCTGTCTGCAG	4832
QY	3725	GGGGCTGAGCGAGGAATCTTTGGAGCCCGAGGAGTCAAGCTGCACCTGAGCAGTGT	3784	Db	21141	GAATGGAAAGCATTTCAAAAACGGAATGGCAGTTGCCAGTTCCTGGNAATCTAATCCTT	21200
Db	20061	AGAGGCTGAGCGAGGAGTACCTGAGCCTGGGAGGCGAAGGTTGCAGTGAAGTACAT	20120	QY	4833	CTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTTACCAGAGTAT-----	4883
QY	3785	TGCGCCACTGCACTCCAGCTGGGTGACAGGACACAGACCTTGCCTCAAAAAATAGAG	3844	Db	21201	GGGTAGAAAAGGTTATTTATCTGCTATGTTAAATTTGATTCTTAAGCAGATCTACTAGGA	21260
Db	20121	COTGCCACTGCAATCCAGCTGGGCAACAGAGTGAGATGCTGTCTCAAAAAACAAAAAC	20180	QY	4884	TTCCCAAAATGCTATTTCAAATTTAGTCTTATCATATGCAAGACACTGTCTCTAGGAGCCAG	4943
QY	3845	AAAAATTAATAATAATGGAACAACTACAAAGAGC-----TGTTGTCC	3888	Db	21261	TTTCAGCAGTTATCTGAGCATCGTGTATTATTAACAATTTATTGTAAGTTAGGTGCTTT	21320
Db	20181	AAAAAGGTTATAGCATAGCAGGCTTTATGATATGATATAAGAAATGAGTTTATTAT	20240	QY	4944	AAAAACAAGAGGAGGAGAAATCAGTCATTA-----TGTTGGGAA	4981
QY	3889	TAGATGAGTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAAGGCTGT	3948	Db	21321	CACGTATTTGTTTGTTCATTTATTTATGAGCACTGCCATGTCTCAGGCTCTCTGTGGG	21380
Db	20241	TTTCATGCTCCATCTTAATTAATATCATAAATTTGATAAGGCCCTAGTGAATTTTATGTTCA	20300	QY	4982	CAACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAAGCAGCAGAGTACAAAATC	5041
QY	3949	CACCTGCACATATTAATAATATCAATTTCTCAATGTATATCCACACAAAGACTGGTAC	4008	Db	21381	TCAGGTGGAAAGAGATAGACCTGGTCTCTACTGTTGTTATGCTCTAGTTAGTGGATG	21440
Db	20301	AGACTTAATGGTTTTTTTCTTTTAAAAATACATTTGAAATGAGGCTTTTCATATAC	20360	QY	5042	ACACATGCAATCAGTATATATCCAAATCATGTAATATGTGCCTGTAGAGAAGACTAGAGGA	5101
QY	4009	GTGAATGTTTCATAGTAC-CTTTATTACAAAAACCCCAAAAGTAGAGACTATCCAAATATCC	4067	Db	21441	CTTGATAGTAGAAAATCAGTTTATGCAATTAGTTTCTCTGTGGTGGTCTGTGATTGGA	21500
Db	20361	TATACCTTTTATATCACGAATGATGTTAAATTAATAGCACTTACAGTATAAAGTTAC	20420	QY	5102	ATAAACACAAAGAAATCTTAACAGTCATTGTCATTAGACACTAAAGTCTTAATTTATTATTA	5161
QY	4068	ATCAACAAGTGAACAAATTAACAAATGTGCTATATCCATGCAATGGAATACACACCTGC	4127	Db	21501	AGCTTGTGTATACCTATTCAAAAGATTTTAAACATTAGACTTTTCAGTTGATTGCTTATA	21560
Db	20421	ATTTTTCATATAGTGTCTAAACACCAATTTCCATTTTACATTTGCTTAGTTTATCTCCAGTA	20480	QY	5162	GACACTATGATATTTGAGATTTAAAAAATCTTTTAATATTTTAAAAATTTAGAGCTCTCTA	5221
QY	4128	AGTACAAAGGAAGAAC- ---TACTTGGGGATGAATCCCAAGTCAATGACGCTAAATGAA	4183	Db	21561	ATCTTCTCTCTGCTGCTATTTCAGCATTTAGCTGTCACTTTAATCTGTAGAGAGTTTGTCCCA	21620
Db	20481	TAGTTAAAGATTAAGCCATGTAACTTGACAGGTAAATATTAACATGATTAGGAATATA	20540	QY	5222	TTTTTCCATAGTATTCAGTTTGACAAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTT	5281
QY	4184	AGAGTCAGATGAAGGAGGAGATAATGTATGCCATACGAAAT-TCTAGAAAATGAAAT	4242	Db	21621	GTGAGACTGTGATATTAGGTGCTAAATTTCCCTGAGGTTCTCTCTCTCTCTCTTTTTTT	21680
Db	20541	AATGTCTCCACAAGGAAACTGAAACTTATTGCTTGATAAGTGTCTTTGGTTTATAAA	20600	QY	5282	TTTTTTTTTTTTGAGATGGAGTTTTTGGTCTTGTGTGCCATGCTGGAGTGGAGTGGCATGA	5341
QY	4243	AACCTATAGTTACAGAAAGCAATCAGGCGAGCAGCATAGAGCTCACACCTGTAATCCCAG	4302	Db	21681	TTTTTTTTTTTTTGAGAGCGGAGTTTCACTC---TTGCCAGGCTGGAGTGCATATGATGCA	21737
Db	20601	AGTAGAAATGAAGTATATATCCTGGAGCATGTGAAATCATTTGACCGCTTTGCACAAAG	20660	QY	5342	YCATAGCTCACTGCACACCTCCACCTCCTGGGTTCAAGCAAGAGCTGTCCGCTCAGCCCTCC	5401
QY	4303	CACTTTGAGAGGCGACGTGGAAAGTTGCTAGAACCTCAGAGTTTCAAGACACGCTGGGC	4362	Db	21738	TTTTCAGCTCACTGCAACACTTGCTCCCGGTTTCAAGTGAATCTCTTGGCTCAGCCCTCC	21797
Db	20661	GCTAAATAGATATTAGTAGAATATTTCAGTAATATAGGACTTGAACCTGCTGTTTTCTC	20720	QY	5402	GGGTAGATGGGATTACAGGCGCCACACACACTCGGCTAAATGTTGTATTTTTTAGTAG	5461
QY	4363	AACAGAGTGAATCCATTCTCCACAAAATGGGAAAAAAGAAAGCAATCAGTGGTTG	4422	Db	21798	GAGTAGTGGGATTACAGGTGCGCCGCCACCCAGCTATTTTTTGTATTTTTTAGTAA	21857
Db	20721	TGTATGTTAAATAAATTTTCAAGAAATACAAGTTTAGCTATGAACAAATTTATGATTT	20780	QY	5462	AGATGGGGTTTTCACCATGTTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG	5512
QY	4423	TCCGTGGGAGGAGGAGTCAAGAGGGAAGA- --AGCTCTGGTGGGTGAGGGTG	4479	Db	21858	AGAGGGGTTTTCACCATGTTTGGCCAGTCTGGTCTCGAACTCCTGACCTCAG	21908
Db	20781	TATTAATAATCAAGAGGACTACTTATCTTCATATTTGAGAAGTTTATATCATTTGTGCT	20840	RESULT 14			
QY	4480	GTGATTCAGGTTCTGTATCTGACTGTGTGAGCAGTTTGGGTGTTTACATCCAAAAATA	4539	ABN90278			
Db	20841	TTTACTTAGTTTTCAGCATATTTGCTTCCCTCATTTTAAAGTAAATTTCTCTAAGCATA	20900	ID	ABN90278	standard; DNA; 28588 BP.	
QY	4540	TCGTAGAAATATGCAATTAATATGAGTGGTGGAGTTTACTGTATGTAATATFACCTCAATG	4599	XX	ABN90278;		
Db	20901	TTCTATGCTTTGTTCTCTCTCTGATTTGATATATTTTTCCTCCTAGAACTCAAAC	20960	AC			
QY	4600	TAAGAAAAATAATGTGAAGAAAGTTTCAATTTCTTCCAGCAAAACGTTATTCAAAT	4659	XX			
Db	20961	ATTGAGCTTTACACCTAGCTTCTGAGAGTCACTGTAATGAAAGAGAACCCAGATATT	21020	XX			
QY	4660	TCTGAGCCCTTTTACTTCGCAATTTCTGCACT-TCTGCCCGTACCATTAGGTGACAG	4718	DT	24-JUL-2002	(first entry)	
Db	21021	TCTTGGCCATACCTCGCTTCAAACTTATCCAGGATGTGTTTCATTTTTCAGCTCTTGAG	21080	DE			
QY	4719	CACTAGCTCCACAAATGGGATAAATGCATTTCTTGGAAAAAGACTTAGGCACAAA- ---T	4772	DE			

Human liver antigen HLDRI94 genomic sequence, SEQ ID NO:399.

Human: liver antigen; liver disorder; hepatic disorder; infection;
hepatitis; viral; parasitic; bacterial; fungal; inflammatory condition;
cirrhosis; granulomatous hepatitis; toxin damage; drug damage;
autoimmune disease; Wilson's disease; primary biliary cirrhosis;
neoplastic disorder; cancer; tumour; portal hypertension;
gastrointestinal disorder; hepatitis; drug screening; gene therapy;
chromosome mapping; forensic analysis; antibody preparation;

Db 20001 ACCAAAATACAAACCAATAGCCAGGTGAGTGGCATGTGCTGGTCCCATAGATATC.20060
 Qy 3725 GGGGGCTGAGCGAGGAGAAATCTTTTGGAGCCAGGAGGTCAAGGCTGCACCTGACGAGTGCT 3784
 Db 20061 AGAGGCTTGAGCGAGGAGGATTAACCTGAGCCTGGGAGCGAAGGTGTGCACTGAGCTGACAT 20120
 Qy 3785 TGGGCCACTGCATCCAGCCTGGGTGACAGGACGAGACCTTGCTCAAAAAATAAGAAG 3844
 Db 20121 CGTGCCACTGCAATCCAGCCTGGGCAACAGAGTGAGATCCTGTCTCAAAAAACAAAAAC 20180
 Qy 3845 AAAAAATTAATAATTAATGGAACAACACTACAAAGAGC-----TGTGTGTC 3888
 Db 20181 AAAAAAGGTTATAGCATAGCAAGGCTTTATGATATGATATAAGAAATGAGTTTTTTTAT 20240
 Qy 3889 TAGATGAGCTACCTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGT 3948
 Db 20241 TTATGTCCATCTTAATTAATAATCAATAATTTGATAAGGCCCTAGTGAATTTTAGTTTCA 20300
 Qy 3949 CAGCTGCACCTACATTAATAATCAATCTCAATGTATATCCACACAAAGACTGGTAC 4008
 Db 20301 AGACTTAATGGTTTTTTTTTTTAAATTTACATTTGAAATGAGGGCTTTTCATATAC 20360
 Qy 4009 GTGAATGTTTCATAGTAC-CTTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCC 4067
 Db 20361 TATACCTTTTATATCAGGAATGCTTAATTTAAATAGCCTTACAGTATAAATCTTAC 20420
 Qy 4068 ATCAACAAGTGAACAATAAACAATGTGCTATATCCATGCAATGGAATACCAACCCCTGC 4127
 Db 20421 ATTTTCAATAGATGTCTAAACACCAATTTCCATTTTACATTTGCTTAGTTTATCTCCAGTA 20480
 Qy 4128 AGTACAAAGGAAGAAGC----TACTTGGGGATGAATCCCAAGCTATGACGCTAAATGAA 4183
 Db 20481 TAGTTAAAGATTAAGCCATTGTAACCTTGACAGTAAATATAAACATGATTAGGAATAAA 20540
 Qy 4184 AGAGTCAGACATGAAGGAGAGATATGATGCCATACGAAAT-TCTAGAAAATGAAAGT 4242
 Db 20541 AATGTCTCCACAAGGAAACTGAACTTATGCTTGATAAGTGCTTGGTTGTTATATAA 20600
 Qy 4243 AACTTATAGTTACAGAAACCAATCAGGCGAGCATAGAGGCTCACACCTGTAATCCGAG 4302
 Db 20601 AGTAGAATGAAGTATATATCCTGGAGGCATGTGAATATCATTTGACCGCTTTGCAACAAG 20660
 Qy 4303 CACTTTGAGAGCCACGTGGGAAGATTGCTAGAACTCAGAGGTTCAAGACCAGCCTGGCG 4362
 Db 20661 GCTAAATAGAATATTAGTAGAATATTCAGTAAATATAGGACTTGAACCTGCTGTTTCTC 20720
 Qy 4363 AACACAGTCAAACTCCATCTCCACAAAATGGGAAAAAAGAAAGCAAACTAGTGGTGG 4422
 Db 20721 TGTATGTTAAAAATAATTTTCAAGAAAAATACAAGGTTAGGCTATGAACAAATTTATGATT 20780
 Qy 4423 TCCTGTGGGGGGGGAAGGACTGCAAGAGGGAAGA---AGCTCTGGTGGGTGAGGGTG 4479
 Db 20781 TATTAATAATCATAGAAAGGACTACTTATCTTCATATTTGAGAAGTTTATATCATTTGCT 20840
 Qy 4480 GTGATTCAGGTTCTGTATCTCTGACTGTGGTAGGAGTTTGGGTGTTTACATCCAAAAATA 4539
 Db 20841 TTTACTTAGTTTTCAGCATATTTGCTCCCTCATTTTAAAGTAAATTTCTTAGCATATA 20900
 Qy 4540 TTCGTAGAATATGCATCTTAATGGGTGGAGTTTACTGTATGTAATATACCTCAATG 4599
 Db 20901 TTTATAGCTTTGTTCTTCTGCTGCTGATTTGATATATTTTCTCTCTAGAACTCAAC 20960
 Qy 4600 TAAGAAAAAATAATGTGTAAGAAAGTTTCAATTTCTTTGCGACGAAACGTTTATTCAAAT 4659
 Db 20961 ATTGAGCTTTACACCTAGCTTCTGAGAGTCACTGTAATGGAAAGAGAGCCAGCATATTT 21020
 Qy 4660 TCCTGAGCCCTTTACTTCGCAAAATCTCTGCACT-TCGTGCCCCGTACCAATTAGTGACAG 4718
 Db 21021 TCTTGGCCATACCTCGCTTCAAACTTATCCAGGATGTGTTCAATTTTGCAGCTCTTGGAG 21080
 Qy 4719 CACTAGCTCCCAAAATTTGGATTAATGCAATTTCTGGAAGAGACTAGGACAAAA-----T 4772
 Db 21081 CAGTTCCATCTCCTATTGATTGTTTACCTCTGCAATTCGAGGCTGAAGACTCTACATGTC 21140

Qy 4773 CCAGGCGATCATCTTGTGCTTTTCATATCAACCACGCTGTACAGCTTGTGCTGTCTGCAG 4832
 Db 21141 GAATGGAAGAGCATTTTCAAAAACGGAATGGCAGTTGCCAGTTTCCCTGGAATCTAACTTT 21200
 Qy 4833 CTGCAATGGGGACTCTTGTATTTCTTTAAAGAAAATTTGGGTACCAGAGTAT----- 4883
 Db 21201 GGGTAGAAAAGGTTATTTATCTCTGGTATGTTAATTTGATTTCTTAAGCAGATCTACTAGGA 21260
 Qy 4884 TTCCACAAATGCTATTTCAAAATTAGTCTTATGATATGCAAGACACTGTCTAGGAGCCAG 4943
 Db 21261 TTTTCAGCAGTATCTCTGAGCATCGTGTATGTAACATTTATTTGTAAGTGTAGTCTCTTT 21320
 Qy 4944 AAAACAAAAGAGGAGGAATCAGTCATTA-----TGTCGGAA 4981
 Db 21321 CAGGTATGTTTTTGTTCATTTTATTTAGCAGCTGCCATGTGTGAGGCTCTGTGGTGG 21380
 Qy 4982 CAACATAGCAAGATATTTAGATCATTTTGCATAGTTTAAAAAGACAGCAGACACAAATC 5041
 Db 21381 TCAGGTGGAAGAGATAGACCTGGTCTCTACTGTGTGTTGTTAGTCTAGTTAGTGGATG 21440
 Qy 5042 ACACATGCAATCAGTATATATCCAAATCATGTAATATGTCCTGTAGAAAGACTAGAGGA 5101
 Db 21441 CTTGGATAGTAGAAAATCAGTTTATGCAATTTAGTTGTTCCCTGTGGTGTCTGTGATTCA 21500
 Qy 5102 ATAAACACAAAGAAATCTTAACAGCTCATTTGATAGACACTTAAGTCTAATTTATTTATTA 5161
 Db 21501 AGGCTTGTCTATACCTATTTCAAGATTTTAAATTTAGACTTTTTCAGTTGATTGCTTATA 21560
 Qy 5162 GACATATGATATTTGAGATTTTAAATAATCTTTAATATTTTAAAAATTTAGAGCTCTTCTA 5221
 Db 21561 ATCTTGTCTGTCTGCTATTTACAGATTTAGCTGTCACTTTAATCTGATAGAGGTTTGTCCCA 21620
 Qy 5222 TTTTTCCTAGTATTTCAAGTTTGAACATGATCAAGTATTAATCTTTCTTTTCTTTTCTTTT 5281
 Db 21621 GTTGAGACTTGATATAGGTGCTAAATTTCCCTGAGGTTCTCTCTCTCTCTCTTTTCTT 21680
 Qy 5282 TTTTCTTTTGTGAGATGGAGTTTGTGCTTTTGTCCCATGCTGGAGTGAATGCATGA 5341
 Db 21681 TTTTCTTTTGTGAGCGAGTTTTCACCTC---TTGCCCGAGGCTGGAGTGAATGATGCA 21737
 Qy 5342 YCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCTGTCCGCTCAGCCTCCC 5401
 Db 21738 TTTTCAGCTCACTGCAACCTCTGCCCTCCCGGTTCAAGTGAATGATGATGCA 21797
 Qy 5402 GGGTAGATGGGATTACAGCGCCGCCACCACTGCGGCTAATGTTGTTGATTTTGTAGTAG 5461
 Db 21798 GAGTAGCTGGGATTACAGGTGCCGCCACCAAGCCAGCTATTTTTTGTATTTTAGTAA 21857
 Qy 5462 AGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCTCAAACTCCTGACCTCAG 5512
 Db 21858 AGCGGGGTTTTCACCATGTTGGCCAGTCTGTGCTCGAACTCCTGACCTCAG 21908

RESULT 15

AA05277 standard; DNA; 10855 BP.

AC AA05277;

DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 7965.

XX Human; reproductive system related antigen; reproductive system disorder; cancer; gene therapy; ds.

OS Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01339.
XX
31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:34:44 ; Search time 7435.16 Seconds
(without alignments)
12010.756 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acacagaatacatgtgcc.....tcaactctgacctcagag 5514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	291.4	5.3	293	17 A0042682	A0042682 CIT-HSP-2
2	199.4	3.6	694	17 AG048607	AG048607 Pan trogl
3	199.2	3.6	373	17 A0045714	A0045714 RPCI11-35
4	199.2	3.6	408	17 A0046391	A0046391 RPCI11-35
5	196.4	3.6	403	13 BG941633	BG941633 ax15d09.x
6	196	3.6	689	14 BM990649	BM990649 UI-H-DIO-

C	7	195.2	3.5	894	14 BQ932123	BQ932123 AGENCOURT
C	8	195.2	3.5	1769	11 AF130079	AF130079 Homo sapi
C	9	194.8	3.5	1027	12 BG539940	BG539940 602567448
C	10	194.4	3.5	1016	14 BQ674146	BQ674146 AGENCOURT
C	11	194.2	3.5	528	17 A0734866	A0734866 HS_3051_A
C	12	194.2	3.5	532	17 A0356384	A0356384 CITBI-EI-
C	13	194	3.5	451	14 BM989249	BM989249 UI-H-DPO-
C	14	194	3.5	885	12 BF575954	BF575954 602132930
C	15	193.4	3.5	778	17 A0743399	A0743399 HS_5387_B
C	16	193.2	3.5	801	9 AUI18632	AUI18632 AUI18632
C	17	192.8	3.5	550	9 AI754064	AI754064 CF19C04.X
C	18	192.8	3.5	916	14 BQ673918	BQ673918 AGENCOURT
C	19	192.2	3.5	529	12 BF897850	BF897850 QV1-MT022
C	20	192	3.5	370	12 BF928499	BF928499 IL2-NT020
C	21	192	3.5	513	14 BM768638	BM768638 K-EST0051
C	22	191.8	3.5	424	17 A0807427	A0807427 HS_3177_A
C	23	191.8	3.5	468	17 A0530817	A0530817 RPCI-11-3
C	24	191.8	3.5	654	14 BM999209	BM999209 UI-H-DIO-
C	25	191.8	3.5	764	12 BG428850	BG428850 602494635
C	26	191.6	3.5	554	9 AL713338	AL713338 DRF2P686P
C	27	191.6	3.5	872	14 BQ674797	BQ674797 AGENCOURT
C	28	191.2	3.5	810	14 BQ670979	BQ670979 AGENCOURT
C	29	191	3.5	440	10 BE677227	BE677227 7d82b10.x
C	30	191	3.5	504	10 AV755268	AV755268 AV755268
C	31	190.8	3.5	402	17 A0087306	A0087306 HS_2191_B
C	32	190.6	3.5	279	14 C15966	C15966 C15966 Clon
C	33	190.6	3.5	425	17 A0513367	A0513367 HS_3073_B
C	34	190.6	3.5	796	17 A0490320	A0490320 RPCI-11-2
C	35	190.4	3.5	654	14 BQ188600	BQ188600 UI-E-EJ1-
C	36	190	3.4	670	17 AG104085	AG104085 Pan trogl
C	37	189.6	3.4	394	10 BE077105	BE077105 RC5-BT060
C	38	189.6	3.4	667	10 AV710119	AV710119 AV710119
C	39	189.6	3.4	672	17 AG102362	AG102362 Pan trogl
C	40	189.4	3.4	380	12 BF803038	BF803038 CMI-C1009
C	41	189.4	3.4	443	9 AA595661	AA595661 n149h02.s
C	42	189.4	3.4	599	9 AL702649	AL702649 DRF2P6860
C	43	189.2	3.4	336	9 AI310670	AI310670 ta46g10.x
C	44	189.2	3.4	826	14 BQ438665	BQ438665 AGENCOURT
C	45	189	3.4	329	9 AI358712	AI358712 qx14c02.x

ALIGNMENTS

RESULT 1	AQ042682	CIT-HSP-2326M11.TR	293 bp	DNA	linear	GSS 14-JUL-1998
AQ042682	LOCUS	DEFINITION	sequence.			genomic clone 2326M11, DNA
ACCESSION	AQ042682	VERSION	AQ042682.1	GI:3310067		
KEYWORDS	GSS.	SOURCE	human.			
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.					
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)					
JOURNAL	Unpublished (1998)					
COMMENT	Other_GSSs: CIT-HSP-2326M11.TV					
	Contact: Mark Adams					
	Department of Eukaryotic Genomics					
	The Institute for Genomic Research					
	9712 Medical Center Dr., Rockville, MD 20850, USA					
	Tel: 301 838 0200					
	Fax: 301 838 0208					
	Email: mdadams@tigr.org					
	Clones are available from Research Genetics (info@resgen.com). BAC					
	end search page:					

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source
1..293
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2326M11"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 69 a 74 c 91 g 59 t

Query Match 5.3%; Score 291.4; DB 17; Length 293;
Best Local Similarity 99.7%; Pred. No. 7.5e-23;
Matches 292; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3535 GACATTTAAAGTGAACAGACAGCCAGGTGTGGTGCCTCAGCGCTGTAAATCCAGCACTC 3594
|||||
Db 1 GACATTTAAAGTGAACAGACAGCCAGGTGTGGTGCCTCAGCGCTGTAAATCCAGCACTC 60
|||||

QY 3595 TGGGAGGTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCGCTGAGCAACA 3654
|||||
Db 61 TGGGAGGTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCGCTGAGCAACA 120
|||||

QY 3655 TGGCAAAACCTGTTTCTATACAAAATAGCCGGCATGTGGCATGTGCTGTGTC 3714
|||||
Db 121 TGGCAAAACCTGTTTCTATACAAAATAGCCGGCATGTGGCATGTGCTGTGTC 180
|||||

QY 3715 CCAGTACTAGGGGCTGAGCAGGAGATCTTTGGAGCCAGGAGGTCAAGGCTGCAC 3774
|||||
Db 181 CCAGTACTAGGGGCTGAGCAGGAGATCTTTGGAGCCAGGAGGTCAAGGCTGCAC 240
|||||

QY 3775 GAGCAGTCTGCGCCACTGCACCTCCAGCTGGGTGACAGCAGCACTTGC 3827
|||||
Db 241 GAGCAGTCTGCGCCACTGCACCTCCAGCTGGGTGACAGCAGCACTTGC 293
|||||

RESULT 2
AG048607 694 bp DNA linear GSS 02-NOV-2001
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-028H18.F, genomic survey sequence.
ACCESSION AG048607
VERSION AG048607.1 GI:16585499
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-028H18.F.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 694)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbeseqsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY

Vector : pKSI45
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..694
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-028H18.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 144 a 206 c 157 g 187 t
ORIGIN

Query Match 3.6%; Score 199.4; DB 17; Length 694;
Best Local Similarity 81.2%; Pred. No. 4.1e-13;
Matches 246; Conservative 0; Mismatches 51; Indels 6; Gaps 1;

QY 3552 AGACAGCCAGGTGTGGTGCCTCAGCGCTTAATCCAGCACTCTGGAGGCTGAGTGGG 3611
|||||
Db 210 AGGTGGCCAGTTTGTGGCTCACACCTGTAATCCAGCACTTTGGGAGGCCGCGG 369
|||||

QY 3612 TGGATCGCTTGAGCCCTGAGTTCAAGACCGCTGAGCAACATGCCAAACCCCTGTTTC 3671
|||||
Db 270 TGGATCACCTGAGGTCAAGGATTCGAGACCGCTGGCAACATGGTGAACCCCTGTTTC 329
|||||

QY 3672 T-----ATACAAAAATTAGCGGGCATGTGGCATGTGGTCCCGCACTACTAG 3725
|||||
Db 330 TACTAAATAACAAAATTAGCCAGCATGTGGTACATGCCTGTAAATCCAGCTACTCG 389
|||||

QY 3726 GGGGCTGAGGAGGAGAAATCTTTGGAGCCAGGAGGTCAAGGCTGCACCTGAGCAGTCT 3785
|||||
Db 390 GGAGCTGAGGAGGAGAAATCACTTGAACCCAGGAGGTGGAGTTGCATTGAGCTGAGATT 449
|||||

QY 3786 GCGCCACTGCATCCAGCTGGGTGACAGCAGCACCTTGCTCAAAAATAAAGACA 3845
|||||
Db 450 GCGCCACTGCATCCAGCTGGGTGACAGCAGCACCTCCATCTCAAAAATAAAGAA 509
|||||

QY 3846 AAA 3848
|||||
Db 510 AAA 512
|||||

RESULT 3
AG045714 373 bp DNA linear GSS 14-APR-1999
LOCUS
DEFINITION RPC111-35A7.TJ RPC1-11 Homo sapiens genomic clone RPCI-11-35A7, DNA
sequence.
ACCESSION AG045714
VERSION AG045714.1 GI:3314641
KEYWORDS GSS.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 373)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
Venter,J.C.
Use of BAC End Sequences for Sequence-Ready Map Building (1998)
Unpublished (1998)
Other_GSSs: RPC111-35A7.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamads@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

TITLE JOURNAL COMMENT

and search page: [search.html](#)

Research Genetics
http://www.tigr.org

FEATURES

```

1. 373
/organism="Homo sapiens"
/gdbid="GDB:7513062"
/db_xref="taxon:9606"
/db_xref="taxon:11-35A7"
/clone="RPC1-11-35A7"
/clone.lib="RPC1-11"
/clone.map="Male"
/sex="Male"
/cell_type="Lymphocytes"
/vector="PBAC3.6"
/notes="Human Male BAC 1
RPC1135A7"

```

[illegible]

```

RESULT 4
A0046391
LOCUS      408 bp      DNA
DEFINITION Homo sapiens genomic clone RPCI-11-35E5, linear
            RPCI-11-35E5, DNA
ACCESSION A0046391
VERSION   A0046391
KEYWORDS  sequence.
SOURCE    human.
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
           1. (bases 1 to 408)
           Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
           Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and
           Venter, J.C.
           Title: BAC End Sequences for Sequence-Ready Map Building (1998)
           Unpublished (1998)
           Other_GSSs: RPCI11-35E5.TK
           Contact: Mark Adams
           Department of Eukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: madams@tigr.org
           Clones are derived from the human BAC library RPCI-11. For BAC
           library availability, please contact pieter de Jong
           (pieter@dejong.med.buffalo.edu). Clones may be purchased from
           BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or
           research Genomics (info@resgen.com). BAC end search page:

```

http://www.tlgf.org/

Location/Qualifiers

FEATURES
SOURCE

```

Loc_408
l_408
/organism="Homo sapiens"
/ab_xref="GDB:751356"
/ab_xref="taxon:9606"
/ab_name="RPC1-11-35E3"
/clone_lib="RPC1-11"
/clone_type="Male"
/cell_type="Lymphocytes"
/site_1="ECORI; Site_1: EcoRI;"
/site_2="ECORI; Site_2: EcoRI;"
/vector="pBACel. Library"
Human Male BAC 149 t

```

67 a 102 c

[illegible]

RESULT 5

BG941633 BG941633 xl Proliferating Human Erythroic Cells
LOCUS ax15d09.cdna clone ax15d09 random, mRNA sequence.
DEFINITION Homo sapiens CDNA clone ax15d09 random, mRNA sequence.

ACCESSION BG941633 GI:14341005

VERSION BG941633

KEYWORDS EST.

SOURCE Homo sapiens
Organism Chordata; Craniata; Vertebrata; Euteleostomi,
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
Taxonomy Eutheria;
Mammalia; 1 (bases 1 to 403)
Author(s) Gubin A.N., Njoroge J.M., Bouffard G.G. and Miller J.L.
Title Expression in proliferating human erythroic cells
Gene expression in proliferating human erythroic cells
Genome database 59 (2), 168-177 (1999)

REFERENCE 9933981 Jeffrey L. Miller
Contact: Laboratory of Chemical Biology
Laboratory of Chemical Diabetes and Digestive and Kidney Diseases
National Institute of Health, Bethesda, MD
Building 10, Room 9B17, National Institutes of Health
USA
20892, USA
Tel: 301 402 2373
Fax: 301 435 5148
Email: jm7f@nih.gov
DNA Sequencing Center (NISC).
Intramural Sequencing Center (NISC).
Plate: 15 row: d column: 09
Seq primer: -21ml3 forward primer (ABI).
Location/Qualifiers
1..403
Zorganism="Homo sapiens"

FEATURES source

```

FEATURES
source
1. .403
/organism="Homo sapiens"
Location,

```

Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
The following repetitive elements were found in this cDNA
sequence: 11-168, >ALU (matched complement) 199-418, >ALU 334-623,
>ALU 634-661, >AT_RichLow_complexity
Seq primer: M13 FORWARD
polyA=Yes.

```

FEATURES .
source
location/Qualifiers
1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5862756"
/clone_lib="NCI_CGAP_D10"
/tissue_type="Lung Focal Fibrosis"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site1: EcoR I; Site2: Not I;
NCI_CGAP_D10 is a cDNA library containing the following
tissue(s): A pool of Lung Focal Fibrosis. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(ATG)18 tail. The sequence tag for this library is
ATACGCGTC."

```

BASE COUNT	ORIGIN	Query Match	Score	DB 14	Length	689
189 a	161 c	170 g	168 t	1	others	
<p> TAG_LI8-UI-H-DIO TAG_TISSUE-lung with fibrosis TAG_SEQ=ATACGGCGTC* </p>						
<p> into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ATACGGCGTC. </p>						

121	GTCTCTCTCAGTAGCTAGCACTTTTAA	180
Db		
3400	GATTAAAAATATGCATTTTAAATTTTAA	3457
Qy		
181	ATTTTAAAAAATAACAGTGGCTGGGT	240
Db		
3460	GTAATACTGGTAATAAATCTAGTTTGC	3511
Qy		
241	GAGGCCGACGCGAGGTGGATCACTTGA	3000
Db		
3520	ATGTACGTGAGGGGAGACATTTAAAGT	3571
Qy		
301	CATAACCT--TGCTCTACTAAAAGTGC	357
Db		
3580	GTAATCCAGCACTCTGGGAGCTGAGGT	3633
Qy		
358	GTAATCCAGCACTTTGGGAGCGGACGAG	417
Db		
3640	CCAGCGCTGAGCAACATGTCAAAACCC	3691
Qy		

Qy	5263	TCCTTCTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGATTTTGTCCTTGTGCCCATG	5322
Db	259	TATTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGACACAGAGATTTACTCTTTGTGCCCAGG	200
Qy	5323	CTGGAGTGGAAATGGCATGAYCATAGCTCACTGCAACCTCCACACTCCCTGGGTCCTCAAGCAA	5382
Db	199	CTGGAGTGCAAATGGCACGATCTCGGCTAAGTGCACCTCCGCTCCCGGGTTCAAGGGAT	140
Qy	5383	GCTGTGCCCTCAGCCTCCCGGTAGATTAGGGATTACAGGCCGCCACCACACACTCGGCTA	5442
Db	139	TCTCCTGCCCTCAGCCTCTCCAGTAGCTGGGGATTACAGGCATGCGCCACCATTGCTGGCTA	80
Qy	5443	ATGTTTCTATTTTATGAPAGATGGGGTTTACCACATGTTGGCCACAGGCTGCTCTCAAACTC	5502
Db	79	ATTTTTGTATTTTATGAPAGATGGGGTTTCCACATGTTGGTCAGGCTGCTCTCAAACTC	20
Qy	5503	CTGACCTCAG	5512
Db	19	CTGACCTCAG	10

RESULT 8	
AF130079	
LOCUS	1769 bp mRNA linear HTC 08-MAY-2001
DEFINITION	Homo sapiens clone FLC0578 PRO2852 mRNA, complete cds.
ACCESSION	AF130079
VERSION	AF130079.1 GI:11493462
KEYWORDS	HTC.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Zhang,C., Yu,Y., Zhang,S., Wei,H., Bi,J., Zhou,G., Dong,C., Zai,Y., Xu,W., Gao,F., Liu,M. and He,F. 1 (bases 1 to 1769)
TITLE	Functional prediction of the coding sequences of 75 new genes deduced by analysis of cDNA clones from human fetal liver
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1769)
AUTHORS	Zhang,C., Yu,Y., Zhang,S., Wei,H., Bi,J., Zhou,G., Dong,C., Zai,Y., Xu,W., Gao,F., Liu,M. and He,F.
TITLE	Direct Submission
JOURNAL	Submitted (23-FEB-1999) Department of Experimental Hematology, Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing, Beijing 100850, P. R. China
FEATURES	Location/Qualifiers
source	1. 1769

```

source
i. .1769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FtC0578"
/tissue_type="liver"
/dev_stage="fetus"
/complement(338. .513)
/rpt_family="L1"
/rpt_type=dispersed
/complement(627. .938)
/rpt_family="Alu"
/rpt_type=dispersed
683. .1192
/notes="predicted protein of H02852"
/codon_start=1
/evidence=not_experimental
/product="PRO2852"
/protein_id="AAC35505.1"
/db_xref="GI:11493463"
/translation="MRENGVILAHCSRLRLSSNPVSASRVAGITGACHHQAQLISV
FLVETGFPHITGAQGLLELTSGDPPAWASASGITVGHSCARLVCACGCIKFGKFFFAAR
HVLIVTVVYLSIMWYISSDIYFYQLLDYKRLRLSCFLIFGVIGPQGPKRRLLDCAI
HVCFHCQDN"
repeat_region
repeat_region
CDS

```

BASE COUNT	INVESTIGATOR	BASE COUNT	INVESTIGATOR
431 a	288 c	347 g	703 t
ORIGIN			

Query Match	3.5%;	Score 195.2;	DB 11;	Length 1769;
Best Local Similarity	80.5%;	Pred. No. 5.9e-13;		
Matches 227;	Conservative 1;	Mismatches 54;	Indels 0;	Gaps 0;
QY 5231	AGTATTCAAGTTTCACAATGATCAAGATATTACTCTCTTCTCTTTTTTTTTTTTTTTTTTTTT	5290		
Db 594	AATTTTCAGCTTCATTCAAGTTTGTTGTTGTTGTTTATTTTATTTTATTTTATTTT	653		
QY 5291	TTTCAGATGGAGTTTTCGTCCTTGTTGGCCATGCTGGAGTGGGAATGGCATGAYCATAGCTC	5350		
Db 654	TTTCAGATGGAGTTTTCGTCCTTGTTGGCCATGGGGAGTGCATGGCGTGATCTTGCGTC	713		
QY 5351	ACTGCAACCTCCACCTCCTCGGTTTCAAGCAAAGCTGTCGCTCAGCCTCCCGGTAGATG	5410		
Db 714	ACTGCAGCCTCCGCGCTCCTGGGTTCACGAATTCCTCCTGTCAGCCTCCCGAGTAGCTG	773		
QY 5411	GGATTACAGGCGCCACACCACTCCGGCTAATGTTTGTATTTTAGTAGAGATGGGGT	5470		
Db 774	GGATTACAGGTGCGCTGCCACAGCCAGCTAATTTCTGTATTTTAGTAGACACAGGGT	833		
QY 5471	TTCAACCATGTTGGCCAGCGCTGGTCTCAAACTCCTGAGCTCAG	5512		
Db 834	TTTCATATATGGTCAGGCTGGTCTTGAACTCCTGACCTCAG	875		

RESULT 9	
LOCUS	BG539940/c
DEFINITION	602567448F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4692322 5', mRNA linear EST 03-APR-2001
ACCESSION	BG539940
VERSION	BG539940.1 GI:13532173
KEYWORDS	mRNA sequence.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 1027)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue procurement: CLONTECH Laboratories, Inc. cDNA Library preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CMI1512 row: 0 column: 11 High quality sequence stop: 322.

```

FEATURES
  source
    high quality sequence stop: 322.
    location/Qualifiers
      1. .1027
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4692322"
        /clone_lib="NIH_MGC_77"
        /lab_host="DH10B (T1 phase-resistant)"
        /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
        Sfil (ggcgctctggcc); site:2: Sfil (ggccattagcc); 5' an
        3' adaptors were used in cloning as follows: 5' adaptor
        sequence: 5'-CAGCGCATATGAGCC-3' and 3' adaptor sequence
        5'-ATTCTAGCCCGAGGGCGGCAGATG-dt(30)BN-3' (where B = A,
        C, or G and N = A, C, G, or T). Average insert size 1.9
        kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
        by PCR. This library was enriched for full-length clones
        and was constructed by Clontech Laboratories (Palo Alto,
        CA). Note: This is a NIH_MGC Library."
        376 a 187 c 280 g 184 t
  BASE COUNT

```

BASE COUNT	376 a	187 c	280 g	184 t
ORIGIN				
Query Match	3.5%; Score 194.8; DB 12; Length 1027;			

Best Local Similarity 86.3%; Pred. No. 9.7e-13;	
Matches 214; Conservative 1; Mismatches 33; Indels 0; Gaps 0;	
QY	5265 TTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATGGAGTTTTCGTCTTGTTGCCCATGCT 5324
Db	309 TTTTCTTTTTCAGACAGAGTTTACTCTGTTGTGCCCAGGCT 250
QY	5325 GGAGTGAANTGGCATGAYCTAGCTACTGCACCTCCACCTCTCGTGGTTCAACAGCAAAGC 5384
Db	249 GGAGTGAANTGGCAGCATCTCGGCTAACTGCAACCTCCGCCTCCCGGGTTCAGCCGATTC 190
QY	5385 TGTCCGCTCAGCCTCCCGGGTAGATGGGATTACAGGGGCCACACACACTCGGCTAAT 5444
Db	189 TCTTGGCTCAGCCTCTCGAGTAGCTGGGATTACAGGCATCGCCACCATGCCTGGCTAAT 130
QY	5445 GTTTTGTATTTTACTAGAGATGGGGTTTCACCATGTTGCCAGGCTGGCTCAAACTCCT 5504
Db	129 TTTTGTATTTTGTAGTAGAGATGGGGTTTCACCATGTTGGTCAGGCTGGCTCAAACTCCT 70
QY	5505 GACCTCAG 5512
Db	69 GACCTCAG 62

RESULT 10	
LOCUS	B0674146
DEFINITION	AGENCOURT_8414352 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6272500
ACCESSION	B0674146
VERSION	5, mRNA sequence.
KEYWORDS	B0674146.1 GI:21784980
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: ARCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCW2450 row: p column: 05 High quality sequence stop: 515.

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Gender</i>	2. <i>Gender</i>
3. <i>Marital Status</i>	3. <i>Marital Status</i>
4. <i>Education</i>	4. <i>Education</i>
5. <i>Income</i>	5. <i>Income</i>
6. <i>Occupation</i>	6. <i>Occupation</i>
7. <i>Religion</i>	7. <i>Religion</i>
8. <i>Political Affiliation</i>	8. <i>Political Affiliation</i>
9. <i>Health Status</i>	9. <i>Health Status</i>
10. <i>Family Size</i>	10. <i>Family Size</i>
11. <i>Home Ownership</i>	11. <i>Home Ownership</i>
12. <i>Travel Frequency</i>	12. <i>Travel Frequency</i>
13. <i>Employment Status</i>	13. <i>Employment Status</i>
14. <i>Spending Habits</i>	14. <i>Spending Habits</i>
15. <i>Life Satisfaction</i>	15. <i>Life Satisfaction</i>
16. <i>Community Involvement</i>	16. <i>Community Involvement</i>
17. <i>Volunteer Work</i>	17. <i>Volunteer Work</i>
18. <i>Charitable Donations</i>	18. <i>Charitable Donations</i>
19. <i>Political Participation</i>	19. <i>Political Participation</i>
20. <i>Environmental Concerns</i>	20. <i>Environmental Concerns</i>
21. <i>Health Insurance</i>	21. <i>Health Insurance</i>
22. <i>Retirement Savings</i>	22. <i>Retirement Savings</i>
23. <i>Real Estate Investment</i>	23. <i>Real Estate Investment</i>
24. <i>Art Collection</i>	24. <i>Art Collection</i>
25. <i>Travel Budget</i>	25. <i>Travel Budget</i>
26. <i>Family Traditions</i>	26. <i>Family Traditions</i>
27. <i>Religious Observance</i>	27. <i>Religious Observance</i>
28. <i>Political Activism</i>	28. <i>Political Activism</i>
29. <i>Environmental Activism</i>	29. <i>Environmental Activism</i>
30. <i>Healthcare Usage</i>	30. <i>Healthcare Usage</i>
31. <i>Family Support System</i>	31. <i>Family Support System</i>
32. <i>Home Renovation</i>	32. <i>Home Renovation</i>
33. <i>Travel Companions</i>	33. <i>Travel Companions</i>
34. <i>Employment Satisfaction</i>	34. <i>Employment Satisfaction</i>
35. <i>Spending Priorities</i>	35. <i>Spending Priorities</i>
36. <i>Life Goals</i>	36. <i>Life Goals</i>
37. <i>Community Engagement</i>	37. <i>Community Engagement</i>
38. <i>Volunteer Hours</i>	38. <i>Volunteer Hours</i>
39. <i>Charitable Giving</i>	39. <i>Charitable Giving</i>
40. <i>Political Views</i>	40. <i>Political Views</i>
41. <i>Environmental Awareness</i>	41. <i>Environmental Awareness</i>
42. <i>Healthcare Access</i>	42. <i>Healthcare Access</i>
43. <i>Family Communication</i>	43. <i>Family Communication</i>
44. <i>Home Maintenance</i>	44. <i>Home Maintenance</i>
45. <i>Travel Itinerary</i>	45. <i>Travel Itinerary</i>
46. <i>Family Values</i>	46. <i>Family Values</i>
47. <i>Religious Beliefs</i>	47. <i>Religious Beliefs</i>
48. <i>Political Engagement</i>	48. <i>Political Engagement</i>
49. <i>Environmental Impact</i>	49. <i>Environmental Impact</i>
50. <i>Healthcare Costs</i>	50. <i>Healthcare Costs</i>
51. <i>Family Dynamics</i>	51. <i>Family Dynamics</i>
52. <i>Home Design</i>	52. <i>Home Design</i>
53. <i>Travel Experiences</i>	53. <i>Travel Experiences</i>
54. <i>Employment Challenges</i>	54. <i>Employment Challenges</i>
55. <i>Spending Patterns</i>	55. <i>Spending Patterns</i>
56. <i>Life Milestones</i>	56. <i>Life Milestones</i>
57. <i>Community Networks</i>	57. <i>Community Networks</i>
58. <i>Volunteer Opportunities</i>	58. <i>Volunteer Opportunities</i>
59. <i>Charitable Causes</i>	59. <i>Charitable Causes</i>
60. <i>Political Movements</i>	60. <i>Political Movements</i>
61. <i>Environmental Policies</i>	61. <i>Environmental Policies</i>
62. <i>Healthcare Options</i>	62. <i>Healthcare Options</i>
63. <i>Family Relationships</i>	63. <i>Family Relationships</i>
64. <i>Home Improvements</i>	64. <i>Home Improvements</i>
65. <i>Travel Plans</i>	65. <i>Travel Plans</i>
66. <i>Family Traditions</i>	66. <i>Family Traditions</i>
67. <i>Religious Practices</i>	67. <i>Religious Practices</i>
68. <i>Political Participation</i>	68. <i>Political Participation</i>
69. <i>Environmental Concerns</i>	69. <i>Environmental Concerns</i>
70. <i>Healthcare Access</i>	70. <i>Healthcare Access</i>
71. <i>Family Communication</i>	71. <i>Family Communication</i>
72. <i>Home Maintenance</i>	72. <i>Home Maintenance</i>
73. <i>Travel Companions</i>	73. <i>Travel Companions</i>
74. <i>Employment Satisfaction</i>	74. <i>Employment Satisfaction</i>
75. <i>Spending Priorities</i>	75. <i>Spending Priorities</i>
76. <i>Life Goals</i>	76. <i>Life Goals</i>
77. <i>Community Engagement</i>	77. <i>Community Engagement</i>
78. <i>Volunteer Hours</i>	78. <i>Volunteer Hours</i>
79. <i>Charitable Giving</i>	79. <i>Charitable Giving</i>
80. <i>Political Views</i>	80. <i>Political Views</i>
81. <i>Environmental Awareness</i>	81. <i>Environmental Awareness</i>
82. <i>Healthcare Costs</i>	82. <i>Healthcare Costs</i>
83. <i>Family Dynamics</i>	83. <i>Family Dynamics</i>
84. <i>Home Design</i>	84. <i>Home Design</i>
85. <i>Travel Experiences</i>	85. <i>Travel Experiences</i>
86. <i>Employment Challenges</i>	86. <i>Employment Challenges</i>
87. <i>Spending Patterns</i>	87. <i>Spending Patterns</i>
88. <i>Life Milestones</i>	88. <i>Life Milestones</i>
89. <i>Community Networks</i>	89. <i>Community Networks</i>
90. <i>Volunteer Opportunities</i>	90. <i>Volunteer Opportunities</i>
91. <i>Charitable Causes</i>	91. <i>Charitable Causes</i>
92. <i>Political Movements</i>	92. <i>Political Movements</i>
93. <i>Environmental Policies</i>	93. <i>Environmental Policies</i>
94. <i>Healthcare Options</i>	94. <i>Healthcare Options</i>
95. <i>Family Relationships</i>	95. <i>Family Relationships</i>
96. <i>Home Improvements</i>	96. <i>Home Improvements</i>
97. <i>Travel Plans</i>	97. <i>Travel Plans</i>
98. <i>Family Traditions</i>	98. <i>Family Traditions</i>
99. <i>Religious Practices</i>	99. <i>Religious Practices</i>
100. <i>Political Participation</i>	100. <i>Political Participation</i>

BASE COUNT
ORIGIN

Query Match 3.5%; Score 194.4; DB 14; Length 1016;
Best Local Similarity 85.0%; Pred. No. 1.1e-12;
Matches 216; Conservative 1; Mismatches 37; Indels 0; Gaps 0;

[illegible]

BASE COUNT	104 a	111 c	82 g	154 t
ORIGIN				

Query Match 3.58; Score 194; DB 14; Length 451;

Qy 5269 TTTTCTTTTTTTTTTTTTTTTGAGATGGAGTTTTTTTTTGGTCTGTTGCCCATGCTGGAG 5328

D_b 170 TTGTAICTTTTTTTTTTTTTTTTGTGACGAGTTTTGCTCTTGTCGCCAGGCTGGAG 229

Qy 5329 TGGAAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCCTGGTTCAAGCAAAAGCTGTC 5388

Db 230 TGCAATGGCGGCATCTGGCTCACTGCACCTCCGGCTTCAAGCGATTCTCCT 289

5389 GCCTCAGCCTCCCGGTAGATGGGATTACAGCGGCCACCACCACTCGGCTAATGTTT 5448

DB	290	GCGCTCAGCCCTCCCGAGTAGCTGGGATTTACAGGSCATGCACACCGCCCGCAGCTAAATTTTTT	349
QV	5449	GTATATTTTTACTATACACATCATCGGGGGTTTTACCATCTCTTTGGGCGCAGGCGCGCTCTCAACATCTCTCACC	5508

350
349
348
347
346
345
344
343
342
341
340
339
338
337
336
335
334
333
332
331
330
329
328
327
326
325
324
323
322
321
320
319
318
317
316
315
314
313
312
311
310
309
308
307
306
305
304
303
302
301
300
299
298
297
296
295
294
293
292
291
290
289
288
287
286
285
284
283
282
281
280
279
278
277
276
275
274
273
272
271
270
269
268
267
266
265
264
263
262
261
260
259
258
257
256
255
254
253
252
251
250
249
248
247
246
245
244
243
242
241
240
239
238
237
236
235
234
233
232
231
230
229
228
227
226
225
224
223
222
221
220
219
218
217
216
215
214
213
212
211
210
209
208
207
206
205
204
203
202
201
200
199
198
197
196
195
194
193
192
191
190
189
188
187
186
185
184
183
182
181
180
179
178
177
176
175
174
173
172
171
170
169
168
167
166
165
164
163
162
161
160
159
158
157
156
155
154
153
152
151
150
149
148
147
146
145
144
143
142
141
140
139
138
137
136
135
134
133
132
131
130
129
128
127
126
125
124
123
122
121
120
119
118
117
116
115
114
113
112
111
110
109
108
107
106
105
104
103
102
101
100
99
98
97
96
95
94
93
92
91
90
89
88
87
86
85
84
83
82
81
80
79
78
77
76
75
74
73
72
71
70
69
68
67
66
65
64
63
62
61
60
59
58
57
56
55
54
53
52
51
50
49
48
47
46
45
44
43
42
41
40
39
38
37
36
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

QY 5509 TCAG 5512

Db 410 TCAG 413

RESULT 14

LOCUS	BF575954	885 bp	mRNA	linear	EST 12-DEC-2000
BF575954/c					

DEFINITION 602132930F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288254 5', mRNA sequence.

ACCESSION BF575954
VERSION BF575954.1 GI:11649666
KEYWORDS EST

REPRODUCED EST.	SOURCE	ORGANISM
	human.	Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 885)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS
REFERENCE

JOURNAL
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg. ph D

contact: robert.bera@clonetechnologies.com
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.

cdNA Library Preparation: CLONETECH Laboratories, Inc.
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the Y.M.A.C.E. Consortium/INT. ST.

round through the I.M.A.G.E. CONSOLIDUM/ELUMU at:
<http://image.llnl.gov>
 Plate: LLCM127 row: k column: 07

High quality sequence stop: 302.
Location/Qualifiers

```
source
1. .885
   /organism="Homo sapiens"
```

```

/db_xref=taxon:9606"
/clone=IMAGE:428254"
/clone_lib=NRU_MGC_81"

```

```
/clone_dir= min_mac_01
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LJB
```

(Clontech); Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggc); 5' and 3' adaptors were used in cloning

as follows: 5' adaptor sequence: 5'-CACGGGCATTATGGCC-3'
and 3' adaptor sequence:
5'-TTCCTGAGCCTGTTTGTGACCACTA-3'

3'-ATTC TAGAGGCGAGGCGCGCGGCAATG-01 (30) BN-3 (where B = A,

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:01:09 ; Search time 231.577 Seconds
(without alignments)
7302.178 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacatagtgcca.....tcaaacctctgacctcgag 5514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	218.6	4.0	4421	2	US-08-257-963B-9
C 2	218.6	4.0	4421	4	US-08-367-841A-9
C 3	218.6	4.0	4421	4	US-08-520-373D-6
C 4	218.6	4.0	4421	5	PCT-US95-07201-9
C 5	199.2	3.6	246240	2	US-08-724-394A-20
C 6	199.2	3.6	246240	2	US-08-724-394A-21
C 7	199.2	3.6	246240	2	US-08-724-394A-22
C 8	195	3.5	17327	1	US-07-906-871-15
C 9	194.8	3.5	72928	3	US-09-009-913-1
C 10	193.8	3.5	43950	4	US-09-735-934A-3
C 11	193.8	3.5	98844	4	US-09-791-211-10
C 12	191	3.5	29629	4	US-09-729-995-3
C 13	190.8	3.5	98844	4	US-09-791-211-10
C 14	190.2	3.4	111282	4	US-09-754-250-3
C 15	188.2	3.4	168575	4	US-09-426-290-1
C 16	188	3.4	2561	4	US-09-270-542-101
C 17	188	3.4	2561	4	US-09-270-542-119
C 18	187.8	3.4	17327	1	US-07-906-871-15
C 19	187.4	3.4	9365	4	US-08-608-285A-8
C 20	187.4	3.4	9365	4	US-09-350-836B-8
C 21	187.4	3.4	9365	4	US-09-370-265-8
C 22	187.4	3.4	14747	4	US-09-608-285A-42
C 23	187.4	3.4	15977	4	US-09-608-285A-59
C 24	187	3.4	15297	4	US-09-817-180-3
C 25	186.6	3.4	112132	4	US-09-741-150-3
C 26	186	3.4	43950	4	US-09-735-934A-3
C 27	184.8	3.4	80246	4	US-09-078-294-4

C 28	184.8	3.4	99500	4	US-09-798-096-10	Sequence 10, Appl
C 29	184.4	3.3	81001	4	US-09-750-580-1	Sequence 1, Appli
C 30	184.4	3.3	81001	4	US-09-750-580-1	Sequence 1, Appli
C 31	184	3.3	72604	4	US-09-268-932-7	Sequence 7, Appli
C 32	184	3.3	72604	4	US-09-657-474-7	Sequence 7, Appli
C 33	183.6	3.3	162450	4	US-09-345-882-1	Sequence 1, Appli
C 34	183.4	3.3	2561	4	US-09-270-542-119	Sequence 101, App
C 35	183.4	3.3	2561	4	US-09-270-542-119	Sequence 119, App
C 36	183.4	3.3	31571	1	US-08-323-443B-1	Sequence 1, Appli
C 37	183.4	3.3	33526	3	US-08-658-136-2	Sequence 2, Appli
C 38	183.4	3.3	33577	3	US-08-658-136-1	Sequence 1, Appli
C 39	183	3.3	35060	3	US-08-814-095-7	Sequence 7, Appli
C 40	182.6	3.3	36159	4	US-09-749-588-3	Sequence 3, Appli
C 41	182.6	3.3	99500	4	US-09-798-096-10	Sequence 10, Appl
C 42	182.4	3.3	14581	4	US-08-520-373D-4	Sequence 4, Appli
C 43	182.4	3.3	14581	4	US-08-520-373D-4	Sequence 4, Appli
C 44	182.4	3.3	22481	4	US-08-367-841A-43	Sequence 43, Appl
C 45	182.4	3.3	22481	4	US-08-367-841A-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-08-257-963B-9/c
; Sequence 9, Application US/08257963B
; Patent No. 5840686
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, S.
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,963B
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEPT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4421 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; FEATURE:
; NAME/KEY: JT101

LOCATION:

```

;
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 7.1 kb Bam HI
; OTHER INFORMATION: fragment Derived from human placental
; OTHER INFORMATION: genomic DNA
;

```

US-08-257-963B-9

Query Match 4.0%; Score 218.6; DB 2; Length 4421;
Best Local Similarity 47.7%; Pred. No. 6.8e-37;
Matches 952; Conservative 1; Mismatches 995; Indels 47;

3538	Qy	ATTTAAAGTGAACACAGACACAGCCAGGTGTGGCTCAGCGCTGTAAATCCAGCACTCTGG	3597
3539			
4107	Db	AGTTAAAGAGGTTTTGCTGGCCAGGTATGGTGGCTCAGCGCTGTAAATCCAGCACTTTGG	4048
3598	Qy	GAGCTCAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACACAGCCTCAGCAACATGG	3657
4047	Db		
3598			
4047	Db	GAGCTAAGGTGGCGGATCACCCTGTCAGAGTTCAAGACACAGCCTGCCAACATGG	3988
3658	Qy	CAAAACCCCTGTTTCT-----ATAACAAAATAATAGCCGGGCATGTGGCATGTGCCCTGT	3710
3987	Db		
3987	Db	TGAACCCCATCTTTACTAAAAATACAAAAAAGTACGCCAGCGTGGTGGCGCACCTGT	3928
3711	Qy	GGTCCAGCTACTAGGS-GGCTGAGCAGAGAAATCTTTGGAGCCCAAGGAGTCAAGGCT	3769
3927	Db	ATCCCCAGCTACTTTGGGAGGCTGAGGCAGAGAAATCTTTGAACCCAGAGGCGGAGGTT	3868
3770	Qy	GCACTGAGCAGTGTCTTGCGCCACTGCACCTCCAGCCTGGGTGACAGGACCAGACCTTGCCT	3829
3867	Db		
3867	Db	GCAGTGAGCGAGATCGCTCCACTGCACCTCCAGCCAGGTGACAGTGTGAGACTCCATCT	3808
3830	Qy	C-AAAAAATAAGAGAATAATTAATAATTAATGGAAACAACACTACAAGAGCTGTTGTCC	3888
3807	Db	CAAAAAAATAAAGAGCGCGGTGGGTCAAGCCCTGTAATCCAGCACTTTGGG	3748
3889	Qy	TAGATGAGCTACTTAGTTAGGCTGATATTTGGTATTTAACTTTTAAAGTCAGGCTCTGT	3948
3747	Db	AGCCGAGGCGCAGATCAGAGGTGAGAGATCCAGACCATCCTGGCTTAACACGGTGA	3688
3949	Qy	CACCTGCACACTATTATTAATAATATCAATTTCTCAATGTATATCCACACAAGAGCTGGTAC	4008
3687	Db	AACCCGCTCTACTAAAAA-----TACAAAAAATTAGTCG	3651
4009	Qy	GTGAATGTTCACTAGTACCTTTATTCAAAAACCCCAAGTAGAGACTATCCAATATCCA	4068
3650	Db	GGCGTGTGGCGACCGCTGTATTCACAGCTACTCGGGAGGTTTGAGCGAGGAATGGG	3591
4069	Qy	TCACAAAGTGAACAAATAAACAATAATGCTATATCCATCGAATGGAATACCACCTGCA	4128
3590	Db	TGAACCCGGAGGAGACTTGCAGTGAGCGGAGATCGCCCACTGCATCCAGCCTGGG	3531
4129	Qy	GTCAAGGAAGAGCTACTTTGGGGATGAATCCCAAGTCAATGACGCTAAATGAAGAGT	4188
3530	Db	TGACAGGCGCAGACTCCATCTCAAGAAAAAATAAATAAAGAGTTCTGCTGCCAG	3471
4189	Qy	CAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTTAGAAATGAAGTAAAC---	4245
3470	Db	ATGTGGGTTCTGTGGGTGGGTGTGGGTCTGTGGGTGGGTGTGGGTCTGCGACCCAG	3411
4246	Qy	--TTATAGTTACAGAAGCAAAATCAGGCGAGGCATAGAGGCTCACACCTGTAATCCCAGC	4303
3410	Db	GTGGGATTCTGTAGTTCTCGAGCCCACTGTAGAGGCTCACACCTGTATCCAGT	3351
4304	Qy	ACTTTGAGAGGCC-ACGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACCCAGCCTGGGC	4362
3350	Db	ACTTTGAGAGCCTGAGGTGGGAGGATCATTGAGCCAGGAGTTCAAAACCCAGCCTGGGC	3291
4363	Qy	AACACAGTGAACCTCCATCTCCACAAAAATGGAAAAAAGAAAGCAATCAGTGGTTG	4422
3290	Db	AATATAGTGAGACCGTATCACTACAAAAAAGTTTGTATTAGTCCCGAGCATGGTGGCAC	3231
4423	Qy	TCCTGTGGGAGGGGAGGACTGCAAGAGGGGAAGAGCTCTGTGGTGGGTGAGGGTGGTG	4482

Db	3230	ATGCCTGTAGTCCCACTACTCGAGAAGCTGAGATAGGGTCACTGAGCCTGGTAGGCTG	3117
Qy	4483	ATTC-----AGGTTCTCTATCCTGACCTGTGGTAGCAGTTTGGGTGTTTACATCCAAAA	4537
Db	3170	AGGCTGCAGTGACGCAAGATCGCACTACTGCACCTCCAGCCTGGGTGACAGATGAGACCC	3111
Qy	4538	TATTTCGTAGAAATTAATGATCTTAATGGGTGGAGTTTACTGTATGTAAATTTATACCTCAA	4597
Db	3110	TGCTCTCAACAAAAAATAACAGGTTCTGTGCTGCTGAGCTTTCGTGAGGAATGAGAC	3051
Qy	4598	TGTAAGAAAAAATAATGTTAAGAAAAAGTTTCAATTCTCTTGGCAGCAAAACGTTATTCAA	4657
Db	3050	TTTCCCACAGCTCGGAAGCCTAGGATATGAGGCTGAGGAGGACTCAACACCGTTGCGATC	2991
Qy	4658	ATTCTCTGAGCCCTTTACTTCGCAAAATTCCTGCACTTCTGCCCGTACCAATTAGGTGACA	4717
Db	2990	AGAGAGACACCTATTTTGGTGAATATTTTGTGAATAGTTAAGTGGCCTCCAGTACAAAC	2931
Qy	4718	GCATGAGTCCACAAATTTGGATAAATGCATTTCTCGAAAAAGACTAGGGACAAATCCAGG	4777
Db	2930	TCCCAGTTCCAGAAAAAGACTTAAAACTAGTATCTGGAGATTCCAAAGTCTTTTAAAGGA	2871
Qy	4778	CATCACTGTGCTTTCATATCAACCAGCTGTACAGCTGTGTTGCTGTCTGCGAGCTGCA	4837
Db	2870	TTTTAGCTGATATGACGAGCATCTGGATGTTAGCTGGGGTTTTGGCTGGATTTCCACCTCG	2811
Qy	4838	ATGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACCAGAGTATTTCCACAAATGCTA	4897
Db	2810	CTCAGTCATGATAGTCTGTTTCCCTGGAACATTCCTGTGGTTTTAGTAGTAATGAGATGT	2751
Qy	4998	TTCAAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAAACAAGAGGAG	4957
Db	2750	GTGCACCACCTCCAGTGGGAACTGAAGCTCTTAGTAGTTACATTTGTAATTAATAAGAAAT	2691
Qy	4958	GAGAAATCAGTCATTATGTGGGAACACATAGCAAGATATTTAGATCATTTTCACCTAGTT	5017
Db	2690	ATAAAACAAGAAATCAGCAAGCCTGCTTACCTCCAGAGACACACCGTTCCACTTACAGTA	2631
Qy	5018	AAAAAGCAGCAGAGTAGTACAAATACACATGCAATCAGTATATATCCAAATCATGTAAATA	5077
Db	2630	TCTAGCTTCTCGACTCCTTTGTTTAAATCGTGGAGTATGAACTGTGACCGAGAGAATTA	2571
Qy	5078	TGTGCCCTGTAGAAAGACTAGAGGAATAAACAACAAGATCTTAACAGCTATTGTCATTAGA	5137
Db	2570	ATATGGAGCCGTTTTATTGAGCAGTTCTCTCCAGGGATGACGGCTGTCAGTGTGTTTCT	2511
Qy	5138	CACTAAGCTCAATTAATTAATTAATTAAGACACTATGATATTTGAGATTTTAAAAAATCTTTAAT	5197
Db	2510	GCCAGTTTTCCAGTCCCTACTCTCTGCCCTTCATCTGTACTCTTGGCAGAGTCAGTAAC	2451
Qy	5198	ATTTTAAATTTAGAGCTCTTCTATTTTCCATAGTATTTCAAGTTTGCATAATGATCAAGT	5257
Db	2450	CAACCCTATAAAATC---AAGTTGTAGTAAGACTTTTATTAACATATGCTCTCCTTC	2394
Qy	5258	ATTACTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGTGGATTTTTGGTCTTTGTTC	5317
Db	2393	TGAGTTCTGTGTTTTGTTGTTTTGTTTTGTTTTTGGAGTGGATTTTCACTCTCTGTTGC	2334
Qy	5318	CCATGCTGGAGTGGAAATGGGATGAYCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAA	5377
Db	2333	CCAGGCTGGAGCGCAATGGCGCATCTGGCTTTACGGCAACCTCCCGGTTTCCGGGTTCAA	2274
Qy	5378	GCAAGCTGTGCTCAGCCTCCCGGTAGATGGGATTACAGCGGCCACCACCAACATC	5437
Db	2273	GTGATTCCTGCTTACGCTCCGAGTAGCTGGGATTTACAGCGGCCGCCACCAAGCCT	2214
Qy	5438	GGCTAATGTTGTAATTTTATAGTAGAGATGGGTTTTACCATTGTTGGCCAGGCTGGTCTCA	5497
Db	2213	GGCTAAT-TTTGTATTTTATAGTAGACGGGATTTTCACTATGTTGGTCAGGTTGGTATCA	2155
Qy	5498	AATCTCTGACCTCAG	5512
Db	2154	AATCCCGGCTCAG	2140

	Query Match	4.0%	Score 218.6;	DB 4;	Length 4421;
	Best Local Similarity	47.7%;	Pred. No. 6.8e-37;		
	Matches 952;	Conservative 1;	Mismatches 995;	Indels 47;	Gaps 9;
QY	3538	ATTAAAGCTGAACACACAGCAGCCAGGTGTGGTGGCTCAGCGCTGTAATCCAGACGACCTCTGG	3597		
Db	4107	AGTTAAAGAGGTTTTCTGCGCCAGGATGTGTGGCTCAGCGCTGTAATCCAGACCTTTGG	4048		
QY	3598	GAGGCTGAGGTGGGTGGATGCTTGGACCCCTGAGCTTCAAGACAGCCGCTGAGCAACATGG	3557		
Db	4047	GAGGCTTAAGGTGGCGGGATGACCTGATGTTCAGGAGTTCAGACAGCGCTGGCCAAACATGG	3988		

Db 2930 TCCAGTCCCGAGAAAGACTTAAACACTAGTATCTGGAGATCCCAAGTCTTTAAAGGA 2871
QY 4778 CATCACTTGTGCTTTCATCAACACGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCA 4837
Db 2870 TTTTACGTATATGACGAGCATCGTATGTACCTGGGGTTTGGCTGAGTCCACCTGG 2811
QY 4838 ATGGGACATCTTATTTTAAAGAACTTGGGTACAGAGTATTTCCACAAATGCTTA 4897
Db 2810 CTCAGTGTATGATGCTGTGTTCCCTGGAACATTCCTGTGTTTGTAGTAGTATGATGT 2751
QY 4898 TTCAAATTAAGTCTTATGATATGCAAGACACTGTGCTAGGAGCCGAGAAACAAAGAGAG 4957
Db 2750 GTGCACCACTCCAGTGGGAACCTGAAGCTCTTAGTAGTTACATGTGAATTTAAAGAAAT 2691
QY 4958 GAGAAATCACTATGTTGGGAACAACATACCAAGATATTTAGATCAATTTTGACTAGTT 5017
Db 2690 ATAAACAGAATAACAGAGCTGTACCTACCTCCAGAGACACACCTTCCACTTACAGTA 2631
QY 5018 AAAAAAGCAGAGTACAAAATCACATGCAATCAATCAATCAATCAATCAATCAATCAAT 5077
Db 2630 TCTAGCTTCTCGACTCCCTTGTAAATCGTGGAGTATGAACCTGTGACACAGAGAATAA 2571
QY 5078 TGTGCTGTAGAAAGACTAGAGATAAACAAGAACTTAAACAGTCAATTTGTCATTAGA 5137
Db 2570 ATATGAGCGGTTTATTGAGCAGTTTCTCCAGGGATGACGGCTGTCACTGTGTCTCT 2511
QY 5138 CACTAAGCTAATTAATTAATTAATGACACTATGATATTTGAGATTTAAATAATCTTTAAT 5197
Db 2510 GCCAGTTTCCAGTCTTACTCTCTGCCCTTCACTCTGTACTCTGTGGCAGAGTCAGTAAC 2451
QY 5198 ATTTTAAATTTAGAGCTCTCTATTTTCCATAGTATTTCAAGTTTGACAAATGATCAAGT 5257
Db 2450 CAACCACTAATAAAATC---AAAGTTGTAGTAAGAACTTTATAACACATATGCCCTTC 2394
QY 5258 ATACTCTTCTTT 5317
Db 2393 TGAGTTGTGGTTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTG 2334
QY 5318 CCATGCTGGAGTGGATGTCATGAYCATAGCTCACTGCAACCTCCACCTCCCTGGGTTCAA 5377
Db 2333 CCAGGCTGAGCGCAATGCGCGATCTTGCTTACGCAACCTCCGCCCTCCGGGTCAA 2274
QY 5378 GCAAGAGCTGCGCTCAGCCTCCCGGGTAGATGGGATACAGCGCCGCCACCAACACATC 5437
Db 2273 GTGATTTCTCTGCTCAGCCTCCGAGTAGCTGGGATTAACAGCGCGCCGCCACCAACG 2214
QY 5438 GGCTAATGTTCTATTTTATAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCA 5497
Db 2213 GGCTAAT-TTTGATTTTATAGAGACGGGATTTCACTATGTTGGTCAGGTTGGTATCA 2155
QY 5498 AACTCCTGACCTCAG 5512
Db 2154 AACTCCCGCCTCAG 2140

RESULT 3
US-08-520-373D-6/c
; Sequence 6, Application US/08520373D
; Patent No. 6451763
; GENERAL INFORMATION:
; APPLICANT: Tombran-Pink, Joyce
; APPLICANT: Steele, Fintan R
; APPLICANT: Chader, Gerald J
; APPLICANT: Becerra, Sofia P
; APPLICANT: Johnson, Lincoln V
; APPLICANT: Rodriguez, Ignacio R
; TITLE OF INVENTION: RETINAL PIGMENTED EPITHELIUM DERIVED NEUROTROPIC FACTOR
; FILE REFERENCE: 2026-4203US1
; CURRENT APPLICATION NUMBER: US/08/520, 373D
; CURRENT FILING DATE: 1995-08-29
; PRIOR APPLICATION NUMBER: 08/377,710
; PRIOR FILING DATE: 1995-01-25

; PRIOR APPLICATION NUMBER: 08/279, 979
; PRIOR FILING DATE: 1994-07-25
; PRIOR APPLICATION NUMBER: 07/894, 215
; PRIOR FILING DATE: 1992-06-04
; PRIOR APPLICATION NUMBER: 07/952, 796
; PRIOR FILING DATE: 1992-09-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 4421
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; OTHER INFORMATION: CDS 66-322
US-08-520-373D-6

Query Match 4.0%; Score 218.6; DB 4; Length 4421;
Best Local Similarity 47.7%; Pred. No. 6.8e-37;
Matches 952; Conservative 1; Mismatches 995; Indels 47; Gaps 9;

QY 3538 ATTTAAAGTGAAACACAGACAGCCAGGTGTGCTGCCTCACGCCCTGTATATCCACGACTCTGG 3597
Db 4107 AGTTAAAGAGGTTTGTCTGCCAGGTATGTGCTCACGCCCTGTATATCCACGACTTTGG 4048
QY 3598 GAGGCTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAACCTGAGCAACATGG 3657
Db 4047 GAGGCTAAGTGGGGGATCACTGATGTCAAGAGTTCAAGACCAACCTGAGCAACATGG 3988
QY 3658 CAAAACCCGTCTCT-----ATACAAAAAATTAGCCGGGATGTCGATGTGCTCT 3710
Db 3987 TGAACCCCATCTTTACTTAAATAACAAAACTAGCCAGGCGTGTGTTGGGCACCTGT 3928
QY 3711 GGTCCAGCTACTTAGG-GGCTGAGCAGGAGAACTCTTTGGAGCCCGAGAGTCAAGGCT 3769
Db 3927 AATCCAGCTACTTTGGAGGCTGAGCAGAGAACTCTTTGAACCCAGCAGGCGGAGTT 3868
QY 3770 GCAGTGAGCAGTGTGTCGCCACTGCATCCAGCCTGGGTGACAGACCAAGACCTTTGCCT 3829
Db 3867 GCAGTGAGCAGAGATCGCTCCACTGCATCCAGCCAGGTGACAGTGTGAGACTCCATCT 3808
QY 3830 C-AAAAAATAAGAGAAAAAATTTAAAAATAATGGAACAACTACAAACAGCTGTTGTCC 3888
Db 3807 CAAAAAATAAGAGGCGCGGTGCGGTGCTCANGCTGTATATCCAGACACTTTGGG 3748
QY 3889 TAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAAGGCTGT 3948
Db 3747 AGCCGAGCAGCAGATCACGAGTCCAGAGATCCAGACCATCTCTGGCTAACACGCTCA 3688
QY 3949 CACCTGCATCATATTATTAATAATATCAATTTCTCAATGTATATCCACACAAAGACTGGTAC 4008
Db 3687 AACCCCGTCTCTACTAAAAA-----TACAAAAAATTAAGTCG 3651
QY 4009 GTGAATGTTTCATAGTACCTTTATTCACAAAAACCCCAAGTAGAGACTATCCAAATATCCA 4068
Db 3650 GCGTGGTGGCGAGCGCTGTATTACAGCTACTCGGAGGTTGAGGAGGAGCAATGGG 3591
QY 4069 TCAACAAGTGAACAAATAAACAATAATGTCTATATCCATGCAATGGAATACCAACCTGCA 4128
Db 3590 TGAACCCGGGAGGAGAGGCTTGCAGTGAGCCGAGATCGGCCACTGCATCTCCAGCCTGG 3531
QY 4129 GTACAAAGGAAGAAGCTACTTGGGGATGAATCCCAAGTTCATGACGCTAAATGAAGAAGT 4188
Db 3530 TGACAGAGCCAGACTCCATCTCAAGAAAAAATAAATAAATAAAGTTCGTCGCCAG 3471
QY 4189 CAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTTCTAGAAATCAAAAGTAAC--- 4245
Db 3470 ATGTTGGTTCTGTGGTGGTGTGCTGTGGTGGTGTGCTGTGGTGTGCTGTGGTGTGCTGT 3411
QY 4246 --TTATAGTTACAGAAAGCAAAATCAGGCGAGGATCAGAGGCTCACACCTGTATATCCAGC 4303
Db 3410 GTGCGGATTTCTGTAGGTTCTGCAGGCGCAGGTAGAGGCTCACACCTGTATATCCAGT 3351
QY 4304 ACTTTGAGAGGCC-ACGTGGGAAGATTGCTAGAACTCAGAGTTCAAGACCAACCTGGCC 4362

Db 3350 ACTTTGAGACGCTGAGGTGGGAGGATCACTTGGACCCAGGAGTTTCAAAACACGACCTGGGC 3291
Qy 4363 AACACAGTGAACCTCCATCTCCACAAAATGGGAAAAAGAAAGCAAAATCACTGGTTG 4422
Db 3290 AATATAGTGAGACGCTATCACTACAAAAGATTTGTTTGTTCACCGAGCATGGGCAC 3231
Qy 4423 TCCTGTGGGGAGGGAAGGACTGCAAGAGAGGAAGAGCTCTGGTGGGGTGGAGGTGGTG 4482
Db 3230 ATGCTGTAGTCCACACCTACTCGAGAGAGCTGAGATAGGTCACCTGAGCCTGGTAGGCTG 3171
Qy 4483 ATTC-----AGTTCTGTATCCTCACTGTGTGTAGCAGTTTGGGGTGTTTATCATCCAAAA 4537
Db 3170 AGGTGCGAGTGAGCAAGATCGCACTACTGCACCTCCAGCCTGGGTGACAGAGTAAGACCC 3111
Qy 4538 TATTCTGTAGATATGTCATCTTAATGGTGGAGCTTACTCTATGTGTAATTTATACCTCAA 4597
Db 3110 TGTCTCAACAAAAATAACAGGTTCTGTGCTGCTGAGCTTTCGTGAGGAAGATAGAC 3051
Qy 4598 TGTAAAGAAAAAATAATGTGTAAGAAAAAGTTTCAATCTCTTCCAGCAAAAGTTTATTCAA 4657
Db 3050 TTTCCCAACAGTCCGAAGCCTAGGATATGAGGGCTGAGGAGACTCAAAACGTTGCGAGTC 2991
Qy 4658 ATTCTCGAGCCTTTACTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATATGAGTGACA 4717
Db 2990 AGAGAGACACCTATTTTGGTGAATATTTTGAATAAGTTAAGTGGGCTCCAGTACAAAC 2931
Qy 4718 GCACCTAGCTCCACAAAATGGATAAATGCATTTCTGGAAAAGACTAGGACAAAATCCAGG 4777
Db 2930 TCCAGTTCCGAGAAAAGACTTAAACTAGTATCTGGAGATTCCTCAAGTCTTTAAAAGGA 2871
Qy 4778 CATCACTGTGCTTTTCATATCAACACGCTGTACAGCTTGTGTCCTCTGTCGACGTGCA 4837
Db 2870 TTTTGTGTATATCAGGCATCTGATGTTAGCCTGGGTTTGGCTGGATTCACCTCG 2811
Qy 4838 ATGGGGACTCTGTGATTTCTTTAAGAAACTTTGAGGTACCAGAGATTTTCCACAAATGCTA 4897
Db 2810 CTCAGTGTATGATCTGTTTCCCTGGAACTTCTCTGTGTTTGTAGTAGTAAGATGT 2751
Qy 4898 TTCAAAATAGTGTATGATATGAAGACACTGTGTAGGAGCCAGAAAACAAAGAGGAG 4957
Db 2750 GTGCACACCTCCAGTGGGAAGTGAAGCTTTAGTAGTTACATTTAATAAAGAAT 2691
Qy 4958 GAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTT 5017
Db 2690 ATAAACAGAAATCAGCAGCCTGCTTACCTCCAGAGACACACCGTTCCACITTAGAGTA 2631
Qy 5018 AAAAAGCAGCAGAGTCAAAAATCACACATGCAATCAGTATPAATCCAAATCATGTAATA 5077
Db 2630 TCTAGCTTCTCGACTCCTTTGTTAAAAATCGTGAGTATGAACGTGACCCAGAGAAGTAA 2571
Qy 5078 TGTCCCTGTAGAAAGACTAGAGGATAAACAACAAGATCTTAACAGTCATTTGTCATTAGA 5137
Db 2570 ATATGGAGCCGTTTTATTGAGCAGTTTCCCTCCAGGGATGAGCGCTGTCAGTGTCTTCC 2511
Qy 5138 CACTAACTCTAATTTATTATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTAAAT 5197
Db 2510 GCCAGTTTCCAGTCCCTTACTCTCGCCCTTCATCTGTACTCTTGGCAGAGTCAGTAAC 2451
Qy 5198 ATTTTAAAAATTTAGAGCTCTTCTATTTTTCATAGTATTTCAAGTTTGACAATGATCAAGT 5257
Db 2450 CAACCACCTAATAAATC---AAAGTTGTAGTAAAGACTTTTATAACATATGCTCCTTC 2394
Qy 5258 ATTACTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGATGGAGTTTGGTCTGTGC 5317
Db 2393 TGAGTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 2334
Qy 5318 CCATGCTGGAGTGAATGGATGATGATGCTCACTGCAACTCCACCTCTCTGGTGTCAA 5377
Db 2333 CCAGGCTGGAGCGCAATGGCGCATCTTGGCTTAGGCAACCTCCGCTTCCGGGTTCAA 2274
Qy 5378 GCAAAAGCTGTGCGCTCAGCCTCCCGGGTGTAGTGGATTACAGGCGCCACACCACTC 5437
_ _ _ _ _

Db 2273 GTGATTCCTCCTGCTTCCAGCCTCCCGAGTAGCTGGGATTTACAGGCGCCGCCACACGCT 2214
Qy 5438 GGCTAATCTGTTGTTATTTTCTAGTAGAGATGGGTTTTCACCATGTTGGCCAGGCTGGTCTCA 5497
Db 2213 GGCTAAT-TTTGTTATTTTCTAGTAGAGCGGATTTCACTA1GTTGGTCAGGTTGGTATCA 2155
Qy 5498 AACTCCTGACCTCAG 5512
Db 2154 AACTCCGCGCTCAG 2140
RESULT 4
PCT-US95-07201-9/c
; Sequence 9, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,841
; FILING DATE: 30-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4421 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; FEATURE:
; NAME/KEY: JTI
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 7.1 kb Bam HI
; OTHER INFORMATION: fragment Derived from human placental
; OTHER INFORMATION: genomic DNA; Also referred to as JTI01
PCT-US95-07201-9
Query Match 4.0%; Score 218.6; DB 5; Length 4421;

Best Local Similarity 47.7%; Pred. No. 6.8e-37; Matches 952; Conservative 1; Mismatches 995; Indels 47; Gaps 9;			
QY	3538	ATTAAAGTGAACACAGCAGCGAGGTGGTGGCTCAGCGCCTGTAATCCAGCACTCTGG	3597
Db	4107	AGTTAAAGAGGTTTGTCTGGCCAGGTATGGTGGCTCAGCGCCTGTAATCCAGCACTTGG	4048
QY	3598	GAGGTGAGGTGGGTGGGATTCGTTGAGCCCTGGAGTTCAAGACCCAGCCTGAGCAACATGG	3657
Db	4047	GAGGTGAGGTGGGTGGGATTCGTTGAGCCCTGGAGTTCAAGACCCAGCCTGAGCAACATGG	3988
QY	3658	CAAAACCCCTGTTCT- - - - - ARAACAAAATAGCCGGCATGGTGGCATGTCCTGT	3710
Db	3987	TGAACCCCACTTCTTAAATAACAAAATAGCCGGCATGGTGGCATGTCCTGT	3928
QY	3711	GGTCCAGCTACTAGGG- -GGCTGAGGCAGGAGATCTTTGGAGCCCGAGGAGTCAAGCT	3769
Db	3927	AATCCAGCTACTTGGGAGGCTGAGCAGGAGATCTTGNACCCAGGAGCGAGGT	3868
QY	3770	GCAGTGAGCGTGTGGCCAGCTGCATCCAGCCTGGGTGACAGGACAGACCTTGGCT	3829
Db	3867	GCAGTGAGCGGAGATCGCTCCAGCTGCATCCAGCCTGGGTGACAGTGTGAGACTCCATCT	3808
QY	3830	C-AAAAAATAGAGAAAATTAATAAATGAACAACTACAAAGAGCTGTGTCC	3888
Db	3807	CAAAAAAATAAGAGCCGGGTGGCTCAAGCCTGTAAATCCAGCACTTTGGG	3748
QY	3889	TAGATGAGCTACTTGTAGGTGATATTTGGTATTTAACTTTTAAAGTCAAGGTCTGT	3948
Db	3747	AGGCGAGCGCAGATCAGGAGTCAGGAGTCCAGACCATCTGTGCTAACACGGTGA	3688
QY	3949	CACCTGCACTACATTTAAATATCAATTCATATATATCCACAAAAGACTGGTAC	4008
Db	3687	AACCCGCTCTACTAAAAA- - - - - TAAAAAATAATAGTCTG	3651
QY	4009	GTGATGTTTCACTAGTACCTTTATTCACAAAACCCAAAGTAGAGACTATCCAAATATCA	4068
Db	3650	GGCTGGTGGCGGAGCCTGTATTCACAGCTACTCGGAGGTGAGGAGGAGAAATGGG	3591
QY	4069	TCAACAAGTGAACAAATAAACAATAATGCCATATATCCATGCAATGGAATACACCCCTGCA	4128
Db	3590	TGAACCCGGAGGAGAGCTTGCAGTGAGCCGAGATCGCCACTGCATCCAGCCTGG	3531
QY	4129	GTACAAAGAGAGAGTACTTGGGATGAATCCCAAGTCAAGCTATGAGCGTAAATGAAGAGT	4188
Db	3530	TCACAGACCCAGACTCCATCTCAAGAAAAAATAATACTAAGAGGTCTGTGCGCAG	3471
QY	4189	CAGACATGAGGAGGAGATATGTATGCCATACGAAATTCAGAAAAATGAAGTAAC- - -	4245
Db	3470	ATGTGGGTCTGTGGGTGGGTGTTCTGTGGGTGGGTGTTCTGCGAGCCAG	3411
QY	4246	- -TTATAGTTACAGAAACAAATCAGGCGAGGCATAGAGGCTCACACCTGTAATCCAGC	4303
Db	3410	GTGCGGATCTGTGTAGTCTTGCAGGCCAGTGTAGAGGCTCACACCTGTAATCCAGT	3351
QY	4304	ACTTTGAGAGGCC- -ACGTGGGAAGATTCGTAGAACTAGGAGTTCAAGACCGCCTGGGC	4362
Db	3350	ACTTTGAGAGCCTGAGGTGGGAGTACCTTGAAGCCAGGAGTTCAAAACCCAGCCTGGC	3291
QY	4363	AACACAGTGAACCTCATCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTG	4422
Db	3290	AATATGTGAGACCGGTATCACTACAAAAGATTTGTTTGTACCCGAGCATGGTGGCAC	3231
QY	4423	TCCTGTGGGAGGGAAGGACTCAAGAGGGAAGAGCTCTGTGGGTGAGGGTGGGTG	4482
Db	3230	ATGCTGTAGTCCCACTACTCGAGAACCTGAGATAGGTCACCTGACCTGGTAGGCTG	3171
QY	4483	ATTC- - - - - AGTTCTGTATCCCTGCTGTGTGTAGCAGTTTGGGTGTTTACATCAAAA	4537
Db	3170	AGGTGCACTGAGCAAGATCCGACTACTGCACTCCAGCCTGGGTGACAGAGTAAGACCC	3111
QY	4538	TATTCGTAGATATGCACTTAAATGGGTGAGTTTACTGTATGTAAATATACCTCAA	4597

Db	3110	TGCTCTCAACAAAAAATAACAGGTTCTGTTGCTGCTGAGCTTTCGTCAGGAAGATAAGAC	3051
QY	4598	TGTAAGAAAAAATAATGTGTAAAGAAAGTTTCAATTTCTTCCAGCAACACGTTATTCAA	4657
Db	3050	TTTCCCAACAGCTCCGAAGCCTAGGATATGAGGGCTGAGGAGGACTCAAACCGTTGCAGTC	2991
QY	4658	ATTCTGTAGCCCTTTTACTTTCGCAAAATCTCTGCACCTTCTGCCCGCTACCATTTAGTGACA	4717
Db	2990	AGAGAGACACCTATTTTGGTGAATATTTTGAATAGTTAAAGTGGCCTCCAGTACAAC	2931
QY	4718	GCACCTAGCTCCACAAATGGGATAAATGCAATTTCTGAAAAAGACTAGGGAACAAATCCAGG	4777
Db	2930	TCCAGTTCACCAAGAAAGACTTAAAACTAGTATCTGGAGATTTCCAGTCCCTTTAAAGGA	2871
QY	4778	CATCACTTGTGCTTTCATATCAACACCGCTGTACAGCTTGTGCTGTCTGCAGCTGCA	4837
Db	2870	TTTTAGCTGATATGAGGCATCTGGATGTTAGCCTGGGTTTTGGCTGGATTTCCACCTCG	2811
QY	4838	ATGGGACTCTTGTATTTCTTTAAGGAAACTTGGGTACCAGAGTATTTCCACAAATGCTA	4897
Db	2810	CTCAGTGATGATGATGCTGTTCCCTGGAAACATTCCTGTGGTTTATAGTAAATGAGATGT	2751
QY	4898	TTCAATTTAGTCTTATGATATGCAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAG	4957
Db	2750	GTGCACCACTCCAGTGGGAACCTGAAAGCTCTTAGTAGTTACATTTGTAATTTAAAAAAGAAT	2691
QY	4958	GAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGTAGCTAGTT	5017
Db	2690	ATAAAACAAGAAATCAGCAAGCCTGTTACCTCCAGAGACACACCGTTCCACTTACAGTA	2631
QY	5018	AAAAAGCAGCAGATACAAAATCACATGCAATCAGTATATTAATCCAAATCATGTAAATA	5077
Db	2630	TCTAGCTTCTCGACTCCTTTGTTAAAAATCGTGGAGTATGAACTGTGACACAGAGTAA	2571
QY	5078	TGTGCTCTGAGAAAGACTAGAGGAATAAACACAAGATCTTAAACAGTCAATTCATATAGA	5137
Db	2570	ATATGGAGCGGTTTATTTAGAGCAGTTTCTCCCGGAGTACCGGCTGTCACTGTGTTCCT	2511
QY	5138	CACTAAGTCTAATTTATTTATTTAGACACTATGATATTTGAGATTTTAAAAATCTTTAAT	5197
Db	2510	GCCAGTTTTCAGTCTTACTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	2451
QY	5198	ATTTTAAATTTAGAGCTCTTCTATTTTCCATATATTTCAAGTATTGACATGATCAAGT	5257
Db	2450	CAACCACTAATAAATC- - - - - AAAGTTGTAGTAAAGACTTTAATACACATATGCCCTTC	2394
QY	5258	ATTACTCTTCTTT	5317
Db	2393	TGAGTTGTGGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTG	2334
QY	5318	CCATCGTGGAGTGGAAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAA	5377
Db	2333	CCAGCTGGAGCGCAATGGCGGATCTTGGCTTACGGCAACCTCCGCCCTCCGGGTTCAA	2274
QY	5378	GCAAGCTGTGCGCTCAGCCTCCCGGTAGATGGATTTACAGGCGGCGCCACACCACTC	5437
Db	2273	GTGATTTCTCTGCTTCAGCCTCCGAGTACGCTGGGATTTACAGGCGGCGCCACCACTC	2214
QY	5438	GGCTAATGTTGTTATTTTGTAGAGATGGGTGTTTCCACATGTTGCCAGGCTGTGTCTCA	5497
Db	2213	GGCTAAT- -TTGTATTTTGTAGTACAGCGGATTTCACTATGTTGGTCAGGTTGGTATCA	2155
QY	5498	AACTCCTGACCTCAG 5512	
Db	2154	AACTCCCGGCTCAG 2140	

RESULT 5
US-08-724-394A-20/C
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.